Gabler EditionWissenschaft

Susanne Hensel-Börner

Validität computergestützter hybrider Conjoint-Analysen

Hensel-Börner
Validität computergestützter hybrider Conjoint-Analysen

GABLER EDITION WISSENSCHAFT

Susanne Hensel-Börner

Validität computergestützter hybrider Conjoint-Analysen

Mit einem Geleitwort
von Prof. Dr. Henrik Sattler

Springer Fachmedien Wiesbaden GmbH

Bibliothek - CIP-Einheitsaufnahme

Susanne:
ıtergestützter hybrider Conjoint-Analysen / Susanne Hensel-Börner.
:itw. von Henrik Sattler. - 1. Aufl..
Dt. Univ.-Verl. ; Wiesbaden : Gabler, 2000
(Gabler Edition Wissenschaft)
Zugl.: Jena, Univ., Diss., 2000

ISBN 978-3-8244-7269-7 ISBN 978-3-663-09092-2 (eBook)
DOI 10.1007/978-3-663-09092-2

1. Auflage November 2000

Alle Rechte vorbehalten

© Springer Fachmedien Wiesbaden 2000
Ursprünglich erschienen bei Betriebswirtschaftlicher Verlag Dr . Th. Gabler GmbH 2000.

Lektorat: Brigitte Siegel / Jutta Hinrichsen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

www.gabler.de
www.duv.de

Höchste inhaltliche und technische Qualität unserer Produkte ist unser Ziel. Bei der Produktion und Verbreitung unserer Werke wollen wir die Umwelt schonen. Dieses Buch ist deshalb auf säurefreiem und chlorfrei gebleichtem Papier gedruckt. Die Einschweißfolie besteht aus Polyethylen und damit aus organischen Grundstoffen, die weder bei der Herstellung noch bei der Verbrennung Schadstoffe freisetzen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Geleitwort

Seit Anfang der 70er Jahre hat die praktische Bedeutung von Conjoint-Analysen als ein Instrument zur Messung von Nachfragerpräferenzen kontinuierlich zugenommen. In den letzten 10 Jahren sind verstärkt Weiterentwicklungen traditioneller Conjoint-Analysen vorgeschlagen worden, insbesondere in Form von computergestützten hybriden Conjoint-Analysen sowie choice-based Conjoint-Analysen. Hybride Conjoint-Analysen verbinden direkte und indirekte Verfahren der Präferenzmessung miteinander. Die Auskunftspersonen geben hierbei zunächst im direkten Befragungsteil (Self-Explicated-Methode) isoliert ein direktes Urteil zur Wichtigkeit von Eigenschaften und deren Ausprägungen ab (z.B. Präferenzurteile über einzelne Produkteigenschaften bzw. deren Ausprägungen) und beurteilen dann im indirekten Befragungsteil ausgewählte Eigenschaftskombinationen bzw. Stimuli (z.B. globale Präferenzurteile hinsichtlich ganzheitlich beschriebener Produkte). Hybride Conjoint-Analysen eignen sich insbesondere dann, wenn eine hohe Anzahl von Eigenschaften auf individueller Ebene zu untersuchen sind. Die meisten praktischen Anwendungen hybrider Conjoint-Analysen hat die Adaptive Conjoint Analysis (ACA) erfahren. Trotz ihrer weiten Verbreitung und der prinzipiellen Vorteilhaftigkeit hybrider Verfahren weist die ACA verschiedene Schwachpunkte auf (z.B. die wiederholte Verwendung von Paarvergleichen).

Vor diesem Hintergrund besteht das Ziel der vorliegenden Arbeit darin, neue computergestützte hybride Conjoint-Analysen zu entwickeln und deren Validität gegenüber bestehenden Ansätzen auf Basis einer breit angelegten empirischen Untersuchung zu analysieren. Die neu entwickelten Conjoint-Analysen beinhalten zum einen (aufbauend auf einem Ansatz von Srinivasan und Park 1997) die Customized Computerized Conjoint Analysis (CCC) und zum anderen die Choice-Oriented Individualized Conjoint Analysis (CHIC). Als Referenzpunkte für die Validitätsprüfung werden die weit verbreitete ACA sowie die Self-Explicated-Methode gewählt. Self-Explicated-Methoden bieten sich als Vergleichsbasis an, da sie prinzipiell für die betrachtete Problemstellung einer individuellen Präferenzmessung bei vielen Eigenschaften geeignet sind und von der Handhabung her Kosten- sowie Zeit- und Komplexitätsvorteile aufweisen.

Die vorgenommenen empirischen Validitätsprüfungen stellen - unterstützt durch das Marktforschungsunternehmen IPSOS Deutschland - eine der umfangreichsten wissenschaftlichen Untersuchungen im Bereich der Conjoint-Analyse dar. Im Einzelnen werden die vier Kriterien diskriminierende Validität, interne Validität, Prognosevalidität und externe Validität überprüft. Hervorzuheben ist die Untersuchung der externen Validität, bei der für jedes Verfahren prognostizierte Marktanteile mit realen Marktanteilen des GfK-Handelspanels verglichen werden.

Die Befunde sind bemerkenswert. Zum einen können die beiden neuen Verfahren CCC und CHIC gegenüber dem bisherigen Standard ACA mindestens ebenbürtige Ergebnisse erzielen. Damit wird die insbesondere in der Praxis weite Verbreitung der ACA in Frage gestellt. Zum anderen weisen die Self-Explicated-Methoden keine gravierenden Nachteile gegenüber sämtlichen hybriden Conjoint-Analysen im Hinblick auf die untersuchten Validitätskriterien auf. Auch wenn sicherlich noch weitere Forschung notwendig ist, so deuten die Befunde darauf hin, daß sich methodischer und datenerhebungstechnischer Aufwand nicht zwingend lohnen muß.

Die Lektüre des Buches kann nicht nur aufgrund dieser interessanten Ergebnisse und der neu entwickelten State-of-the-Art-Conjoint-Analysen empfohlen werden. Darüber hinaus ist das Buch sowohl Wissenschaftlern als auch interessierten Praktikern als ein Einstieg in die Methoden zur Präferenzmessung nahezulegen. Es wird ein sehr gut lesbarer und kompakter, gleichzeitig jedoch umfassender Überblick in die Thematik der Messung von Präferenzen gegeben. Ich bin überzeugt, daß diese Arbeit eine weite Verbreitung in Theorie und Praxis finden wird.

Prof. Dr. Henrik Sattler

Vorwort

Kaum ein anderes Verfahren ist zentraler Gegenstand so vieler Publikationen und Studien wie die Conjoint-Analyse. Fast zu jedem einzelnen Ablaufschritt des Verfahrens liegen zahlreiche Untersuchungen vor, die alternative methodische Vorgehensweisen vorschlagen oder empirisch miteinander vergleichen. Doch trotz dieser umfangreichen und umfassenden Diskussion verschiedener Verfahrensvarianten scheint sich ein Verfahren, die Adaptive Conjoint Analysis (ACA), nahezu als Standardverfahren sowohl in der Wissenschaft als auch in der Praxis zu etablieren. Vor dem Hintergrund der umfassenden Diskussion des methodischen Vorgehens bei der Conjoint-Analyse stellt sich jedoch die Frage, ob sich mit der ACA tatsächlich die geeignetste Verfahrensvariante durchgesetzt hat oder die Dominanz des Verfahrens maßgeblich auf das Vorliegen eines integrierten Softwarepakets zurückgeführt werden muß. Die vorliegende Arbeit soll eine Antwort auf diese Frage geben. Hierfür werden mit der Customized Computerized Conjoint Analysis (CCC) und der Choice-Oriented Indiviualized Conjoint Analysis (CHIC) neue computergestützte Conjoint-Varianten entwickelt und im Rahmen einer empirischen Validitätsprüfung direkt mit der ACA verglichen.

Meinem Doktorvater und akademischen Lehrer, Herrn Prof. Dr. Henrik Sattler, verdanke ich sehr viel. Er hat mir dieses spannende Dissertationsthema vorgeschlagen, war unermüdlich in seiner Diskussionsbereitschaft und hat meine Arbeit und mich persönlich zu jeder Zeit unterstützt. Darüber hinaus hat er es mir ermöglicht, am Aufbau eines neu gegründeten Lehrstuhls an einer sehr jungen Fakultät in den neuen Bundesländern mitzuwirken. Ich konnte dadurch Erfahrungen in vielerlei Hinsicht sammeln, für die ich mindestens genauso dankbar bin. Für die Übernahme des Zweitgutachtens zu meiner Arbeit gilt mein Dank Herrn Prof. Dr. Johannes Ruhland.

Die empirische Überprüfung der entwickelten Conjoint-Verfahren wäre in dem Umfang ohne die Kooperation mit der IPSOS Deutschland nicht möglich gewesen. Für die Durchführung der gesamten Datenerhebung, die Bereitstellung aller Daten und für die sehr gute und unkomplizierte Zusammenarbeit möchte ich

mich insbesondere bei Frau Claudia Keck und Herrn Harald Hasselmann bedanken.

Die Arbeit entstand im Rahmen meiner Tätigkeit am Lehrstuhl für Marketing an der Friedrich-Schiller-Universität Jena. Während dieser Zeit arbeitete ich Tür an Tür mit meinen Kollegen Dr. Grit Zatloukal und Dr. Olaf Ernst. Ihnen möchte ich an dieser Stelle besonderen Dank aussprechen. Mit ihnen habe ich die spannende Aufbauphase eines neu gegründeten Lehrstuhls erlebt und gemeistert. Die gute Atmosphäre an unserem Lehrstuhl wurde dabei wesentlich von Constanze Neukirch geprägt. Auch bei ihr möchte ich mich ganz herzlich bedanken.

In den unterschiedlichen Phasen der Arbeit erfuhr ich Hilfe von zahlreichen Seiten. Allen voran gilt mein Dank Simone Martin, ohne deren nie endende Unterstützung die neuen Verfahren in der Form nicht entwickelt worden wären. Ich habe heute noch ein schlechtes Gewissen, weil sie für die Fertigstellung der Programmierarbeiten auf einen bereits gebuchten Surf-Urlaub verzichtet hat. Gemeinsam mit Thomas Niekler habe ich die Verfahren in einem umfangreichen Pretest auf ihre Anwendbarkeit getestet. Auch bei ihm möchte ich mich bedanken. Frau Colette Friedrich hat mir immer wieder geholfen, die Auswertung der individualisierten Datensätze zu automatisieren. Ohne sie hätten so manche Auswertungsschritte um ein vielfaches länger gedauert. Die ersten Versionen der Arbeit wurden mehrfach Korrektur gelesen. Hierfür möchte ich mich bei Dr. Kathrin Börner, Dipl.-Volksw. Ines Müller, Prof. Dr. Maria Rumpf und Dr. Michel Clement ganz herzlich bedanken. Sie alle haben durch ihre kritische Durchsicht und konstruktiven Hinweise zur Verbesserung der Arbeit beigetragen.

Diese Arbeit wäre nie entstanden, wenn mich meine Eltern nicht stets darin gestärkt und ermutigt hätten, meinen Weg zu gehen. Mein Mann Niels war immer für mich da. Er hat alle Höhen und Tiefen der Arbeit miterlebt und mich auch des öfteren daran erinnert, daß sich die Welt nicht nur um die Conjoint-Analyse dreht. Ihm und meinen Eltern danke ich von ganzem Herzen und widme ihnen diese Arbeit.

Susanne Hensel-Börner

Inhaltsverzeichnis

Geleitwort V

Vorwort VII

Abbildungsverzeichnis XIII

Tabellenverzeichnis XV

Abkürzungsverzeichnis XVII

Symbolverzeichnis XIX

1 Problemstellung und Aufbau der Arbeit 1

1.1 Problemstellung 1

1.2 Ziele der Arbeit 7

1.3 Aufbau der Arbeit 9

2 Ansätze zur Messung von Nachfragerpräferenzen 13

2.1 Kompositionelle Verfahren (Self-Explicated-Methode) 15

2.2 Dekompositionelle Verfahren (Conjoint-Analyse) 19

2.2.1 Systematisierung dekompositioneller Ansätze zur Präferenzmessung 20

2.2.1.1 Ansätze mit simultaner Bewertung aller Stimuli 22

2.2.1.2 Ansätze mit simultaner Bewertung einer Teilmenge der Stimuli 26

2.2.1.3 Ansätze mit simultaner Bewertung von genau zwei Stimuli 28

2.2.2 Empirische Vergleiche zwischen unterschiedlichen dekompositionellen Ansätzen 29

2.2.2.1 Gütemaße 30

2.2.2.2 Bisherige empirische Vergleiche 32

2.3 Vergleich zwischen der Self-Explicated-Methode und der Conjoint-Analyse 37

2.3.1 Theoretischer Vergleich 38

2.3.2 Bisherige empirische Vergleiche 43

2.4 Hybride Verfahren ... 51
2.4.1 Nicht adaptive hybride Verfahren ... 52
2.4.2 Adaptive hybride Ansätze ... 54
2.4.2.1 Nicht computergestützte Verfahren ... 54
2.4.2.2 Computergestützte Verfahren ... 55

2.5 Vergleich hybrider Verfahren mit anderen Ansätzen zur Präferenzmessung ... 57
2.5.1 Bisherige empirische Vergleiche zwischen hybriden Verfahren und der rein dekompositionellen Conjoint-Analyse ... 57
2.5.2 Bisherige empirische Vergleiche zwischen hybriden Verfahren und der Self-Explicated-Methode ... 62

2.6 Zusammenfassung ... 65

3 Untersuchte computergestützte hybride Verfahren ... 69

3.1 Customized Computerized Conjoint Analysis (CCC) ... 70
3.1.1 Self-Explicated-Phase ... 71
3.1.1.1 Datenerhebung ... 71
3.1.1.2 Schätzung der Teilnutzenwerte ... 73
3.1.2 Conjoint-Phase ... 75
3.1.2.1 Anzahl simultan zu bewertender Stimuli ... 75
3.1.2.2 Form der Beurteilungsaufgabe ... 77
3.1.2.3 Schätzung der Teilnutzenwerte ... 78
3.1.3 Kombinationsphase ... 78

3.2 Choice-Oriented Individualized Conjoint Analysis (CHIC) ... 83
3.2.1 Self-Explicated-Phase ... 84
3.2.2 Conjoint-Phase ... 84
3.2.2.1 Anzahl simultan zu beurteilender Stimuli ... 84
3.2.2.2 Form der Beurteilungsaufgabe ... 88
3.2.2.3 Schätzung der Teilnutzenwerte ... 92
3.2.3 Kombinationsphase ... 95

3.3 Adaptive Conjoint Analysis (ACA) 96
3.3.1 Self-Explicated-Phase 97
3.3.1.1 Datenerhebung 97
3.3.1.2 Schätzung der Teilnutzenwerte 99
3.3.2 Conjoint-Phase 100
3.3.3 Kombinationsphase 102

4 Untersuchungsdesign 109

4.1 Untersuchungsobjekt 109

4.2 Stichprobe 111

4.3 Interviewablauf 115
4.3.1 Self-Explicated-Phase 115
4.3.2 Conjoint-Phase 116
4.3.3 Kombinationsphase 122
4.3.4 Zusatzdaten und Kaufsimulation 126
4.3.5 EDV-Implementierung 128

5 Empirischer Validitätsvergleich 131

5.1 Diskriminierende Validität 132
5.1.1 Diskriminierende Validität zwischen alternativen Formen der Präferenzabfrage 133
5.1.2 Diskriminierende Validität zwischen den untersuchten Methoden zur Präferenzmessung 137
5.1.3 Zusammenfassung der Ergebnisse hinsichtlich der diskriminierenden Validität 145

5.2 Interne Validität 147

5.3 Prognosevalidität 152
5.3.1 First-Hit-Rate 153
5.3.2 Individuelle Kendalls τ-Werte 156

5.3.3 Simulierte Marktanteile 159
5.3.3.1 Kaufverhaltensannahmen 160
5.3.3.2 Maximale Prognosevalidität der verschiedenen Methoden 164
5.3.3.3 Einfluß der Kaufverhaltensannahme auf die Prognosevalidität. 168
5.3.4 Zusammenfassung der Ergebnisse hinsichtlich der Prognosevalidität.... 174

5.4 Externe Validität 176
5.4.1 Marktanteilsmodell 178
5.4.2 Maximale externe Validität der verschiedenen Methoden 179
5.4.3 Einfluß der Kaufverhaltensannahme auf die externe Validität 181
5.4.4 Zusammenfassung der Ergebnisse hinsichtlich der externen Validität... 184

5.5 Diskussion der empirischen Ergebnisse 186

6 Fazit der Arbeit 191

7 Anhang 195

7.1 Bildschirmausdrucke des CCC-Interviews 195

7.2 Bildschirmausdrucke des CHIC-Interviews 200

Literaturverzeichnis 205

Abbildungsverzeichnis

Abbildung 1-1: Ansätze zur Präferenzmessung 4
Abbildung 1-2: Aufbau der Arbeit 11
Abbildung 2-1: Überblick über den Aufbau von Kapitel 2 14
Abbildung 2-2: Ablaufschritte kompositioneller Verfahren zur Präferenzmessung 18
Abbildung 2-3: Systematisierung dekompositioneller Ansätze 22
Abbildung 2-4: Systematisierung hybrider Ansätze 52
Abbildung 2-5: Zusammenhang zwischen den dekompositionellen und computergestützten hybriden Verfahren 56
Abbildung 2-6: Überblick über alle diskutierten Ansätze zur Präferenzmessung 67
Abbildung 3-1: Interviewablauf der untersuchten computergestützten hybriden Verfahren im Vergleich 70
Abbildung 3-2: Interviewablauf der CCC 83
Abbildung 3-3: Beispiel zur Umskalierung der Präferenzwerte in der Conjoint-Phase der CHIC 94
Abbildung 3-4: Interviewablauf der CHIC 96
Abbildung 3-5: Interviewablauf der ACA 105
Abbildung 3-6: Merkmale der untersuchten computergestützten hybriden Verfahren im Vergleich 106
Abbildung 5-1: Häufigkeitsverteilung der individuellen mittleren absoluten Abweichungen zwischen den Bedeutungsgewichten auf Basis der Rangdaten und der Kaufwahrscheinlichkeiten (n=242) 135

Tabellenverzeichnis

Tabelle 2-1: Empirische Vergleichsstudien dekompositioneller Verfahren zur Präferenzmessung 33

Tabelle 2-2: Theoretischer Vergleich zwischen der Conjoint-Analyse und der Self-Explicated-Methode 39

Tabelle 2-3: Empirische Vergleiche zwischen der Conjoint-Analyse und der Self-Explicated-Methode 45

Tabelle 2-4: Empirische Vergleiche zwischen hybriden Verfahren und der Conjoint-Analyse 59

Tabelle 2-5: Empirische Vergleiche zwischen hybriden Verfahren und der Self-Explicated-Methode 63

Tabelle 4-1: Untersuchte Eigenschaften und deren Ausprägungen 111

Tabelle 4-2: Umfang und demografische Struktur der Teilstichproben 113

Tabelle 4-3: Testergebnisse der Überprüfung der Teilstichproben auf Strukturgleichheit 114

Tabelle 4-4: Häufigkeitsverteilung der optimalen Gewichtungsfaktoren w_i* im Vergleich 125

Tabelle 4-5: Produktbeschreibungen im Holdout-Set für die simulierte Kaufentscheidung 128

Tabelle 5-1: Unterschiede zwischen den Bedeutungsgewichten auf Basis der Ranking- und den Kaufwahrscheinlichkeitsdaten bei der CCC (Wilcoxon-Test für abhängige Stichproben) 136

Tabelle 5-2: Mittelwerte (Standardabweichungen) der Bedeutungsgewichte der Eigenschaften bei unterschiedlichen Ansätzen zur Präferenzmessung 139

Tabelle 5-3: Nicht signifikante Unterschiede zwischen den Verteilungen der Bedeutungsgewichte der verschiedenen Methoden zur Präferenzmessung (Kolmogoroff-Smirnoff-Test $p>0,1$) 141

Tabelle 5-4: Korrelationskoeffizienten nach Pearson zwischen den Mittelwerten der Bedeutungsgewichte 143

Tabelle 5-5: Mittelwerte (MW) und Standardabweichungen (SA) der Korrelationskoeffizienten zwischen prognostizierten und beobachteten Werten des Kalibrierungssets 149

Tabelle 5-6: Umfang der Teilstichproben nach Eliminierung von Ausreißern 151

Tabelle 5-7: Ergebnisse der Prognosevalidität, gemessen durch *die First-Hit-Rate* *154*

Tabelle 5-8: Irrtumswahrscheinlichkeiten p der χ^2-Tests auf Signifikanz der Unterschiede zwischen der First-Hit-Rate 155

Tabelle 5-9: Mittelwerte und Standardabweichungen der ermittelten Korrelationskoeffizienten (Kendalls τ) zwischen beobachteter und prognostizierter Rangfolge des Holdout-Sets 157

Tabelle 5-10: Ergebnisse der Kolmogoroff-Smirnoff-Tests auf Unterschiede in den Verteilungen der individuellen Kendalls τ-Werte für das Holdout-Set 158

Tabelle 5-11: Maximale Übereinstimmungen zwischen beobachteten und prognostizierten Werten der simulierten Marktanteile gemessen durch Pearson und mittlere absolute Abweichung (MAA) 165

Tabelle 5-12: Beobachtete Marktanteile (in %) im Rahmen der simulierten Kaufentscheidung 168

Tabelle 5-13: Korrelationskoeffizienten (Pearson) zwischen beobachteten und prognostizierten Werten der simulierten Marktanteile bei unterschiedlichen Kaufverhaltensmodellen 170

Tabelle 5-14: Mittlere absolute Abweichungen (MAA in %) zwischen den beobachteten und prognostizierten Werten der simulierten Marktanteile bei unterschiedlichen Kaufverhaltensmodellen.... 171

Tabelle 5-15: Maximale Übereinstimmungen zwischen beobachteten und prognostizierten Werten der realen Marktanteile gemessen durch Pearson und mittlere absolute Abweichung (MAA)..... 180

Tabelle 5-16: Korrelationskoeffizienten (Pearson) zwischen beobachteten und prognostizierten Werten der realen Marktanteile bei unterschiedlichen Kaufverhaltensmodellen 182

Tabelle 5-17: Mittlere absolute Abweichungen (MAA in %) zwischen den beobachteten und prognostizierten Werten der realen Marktanteile bei unterschiedlichen Kaufverhaltensmodellen.... 183

Tabelle 5-18: Gegenüberstellung der Ergebnisse aller betrachteten Gütemaße 187

Abkürzungsverzeichnis

ACA	Adaptive Conjoint Analysis
BTL	Bradley-Terry-Luce-Model
bzw.	beziehungsweise
CA	Conjoint-Analyse
CBC	Choice-Based Conjoint Analysis
CCA	Customized Conjoint Analysis
CCC	Customized Computerized Conjoint Analysis
CC-SE	Conjunctive-Compensatory Self-Explicated-Methode
CHIC	Choice-Oriented Individualized Conjoint Analysis
DM	Deutsche Mark
et al.	et alii (und andere)
f.	und folgende Seite
ff.	und folgende Seiten
FC	First-Choice-Model
ggf.	gegebenenfalls
GfK	Gesellschaft für Konsumforschung, Nürnberg
Hrsg.	Herausgeber
IIA	Independence from Irrelevant Alternatives
Jg.	Jahrgang
k.A.	keine Aussage möglich
kWh	Kilowattstunde
LINMAP	Linear Programming Techniques for Multidimensional Analysis of Preferences
MAA	Mittlere absolute Abweichung
MW	Mittelwert
No.	Number
Nr.	Nummer
n. s.	nicht signifikant
OLS	Ordinary Least Squares
S.	Seite
SA	Standardabweichung

SE	Self-Explicated-Methode
SE-ACA	Self-Explicated-Methode der Adaptive Conjoint Analysis
SE-CCC	Self-Explicated-Methode der Customized Computerized Conjoint Analysis
SE-CHIC	Self-Explicated-Methode der Choice-Oriented Individualized Conjoint Analysis
sig.	signifikant
sog.	sogenannt
u.a.	unter anderem
vgl.	vergleiche
z.B.	zum Beispiel

Symbolverzeichnis

α_{ijk}:	Bewertung für die k-te Ausprägung der j-ten Eigenschaft durch den i-ten Probanden
β_{ijk}:	Teilnutzenwert für die k-te Ausprägung der j-ten Eigenschaft für den i-ten Probanden
ε_{im}:	Störterm
a_i:	Konstantes Glied der Schätzgleichung für den i-ten Probanden
b_i:	Regressionskoeffizient für die Gewichtung der Teilnutzenwerte aus der Paarvergleichsaufgabe für den i-ten Probanden
BG_{im}:	Bekanntheitsgrad des m-ten Produkts beim i-ten Probanden
c_i:	Regressionskoeffizient für die Gewichtung der Teilnutzenwerte aus der Self-Explicated-Phase für den i-ten Probanden
C_i :	Indexmenge der Kernattribute beim i-ten Probanden
D_{ijk}:	Präferenzwert der k-ten Ausprägung bei der j-ten Eigenschaft beim i-ten Probanden
DG_m:	Distributionsgrad des m-ten Produkts
ER_{im}:	Erstkaufrate für das m-te Produkt des i-ten Probanden
G_{ij}:	Umskalierte Wichtigkeit der j-ten Eigenschaft beim i-ten Probanden
I:	Indexmenge der Probanden
J:	Indexmenge der Eigenschaften
K_j:	Indexmenge der Eigenschaftsausprägungen der j-ten Eigenschaft
KW_i :	Prognostizierte Kaufwahrscheinlichkeit für ein Produkt im Validation-Set durch den i-ten Probanden
M:	Indexmenge der Alternativen
MA_m:	Prognostizierter Marktanteil für das m-te Produkt
N:	Stichprobenumfang

NC_i:	Indexmenge der Nicht-Kernattribute beim i-ten Probanden
p_{im}:	Kaufwahrscheinlichkeit für die m-te Alternative für den i-ten Probanden
P_{ijk}^{CA}:	Umskalierter Conjoint-Teilnutzen der k-ten Ausprägung bei der j-ten Kerneigenschaft für den i-ten Probanden
P_{ijk}^{Pairs}:	Aktualisierter Teilnutzen nach dem letzten Paarvergleich der k-ten Ausprägung bei der j-ten Eigenschaft für den i-ten Probanden
P_{ijk}^{SE}:	Self-Explicated-Teilnutzen der k-ten Ausprägung bei der j-ten Kerneigenschaft für den i-ten Probanden
P_{ijk}^{w}:	Gewichteter Teilnutzen der k-ten Ausprägung bei der j-ten Kerneigenschaft für den i-ten Probanden
U_{im}:	Geschätzter Gesamtnutzen für die m-te Alternative für den i-ten Probanden
w_i:	Gewichtungsfaktor beim i-ten Probanden
w_i^*:	Optimaler Gewichtungsfaktor zwischen Conjoint-Teilnutzen und Self-Explicated-Teilnutzen bei den Kernattributen für den i-ten Probanden
x_{jkm}:	k-te Ausprägung der j-ten Eigenschaft bei der m-ten Alternative

1 Problemstellung und Aufbau der Arbeit

1.1 Problemstellung

Mit jeder Kaufentscheidung treffen Konsumenten die Wahl für oder gegen ein bestimmtes Produkt. Auch wenn das Produkt damit letztendlich als Ganzes beurteilt wird, hängt diese Kaufentscheidung maßgeblich davon ab, welche Eigenschaften bei diesem Produkt in welcher Ausprägung auftreten. Damit wird ein Produkt nach Brockhoff (1999) als Bündel von Eigenschaften definiert, deren Kombination vom Anbieter des Produkts vorgenommen wird. Diese Definition beschränkt sich dabei nicht nur auf die physischen Eigenschaften eines Produkts, sondern kann auch auf Pakete von Produkten oder Dienstleistungen erweitert werden.[1] So hängt beispielsweise die Entscheidung für ein Mobilfunkangebot nicht nur von den Eigenschaften des Endgerätes ab, sondern auch von den Tarifstrukturen und Vertragsbedingungen des jeweiligen Netzbetreibers. Die Kaufentscheidung der Konsumenten liefert für den Anbieter allerdings erst nachträglich die Information, ob seine vorgenommene Kombination der Produkteigenschaften den Anforderungen der Kunden entspricht oder nicht. Eine solch verspätete Information über den Erfolg eines Produkts ist vor allem im Fall eines Mißerfolges mit hohen finanziellen Verlusten verbunden und kann die Wettbewerbsfähigkeit eines Unternehmens gefährden.

Für eine erfolgreiche Kombination der Produkteigenschaften ist Voraussetzung, daß den verantwortlichen Produkt-Managern bereits vor der Produkteinführung Informationen über die Bedeutung der einzelnen Eigenschaften eines Produkts für die Konsumenten vorliegen. Denn erst wenn bekannt ist, welche und in welchem Ausmaß einzelne Produkteigenschaften bei der Kaufentscheidung wichtig sind, können die Produkte und Angebote darauf ausgerichtet und der Produkterfolg entscheidend beeinflußt werden. Beispielsweise wird sich ein Kunde, für den die mobile Erreichbarkeit im gesamten Bundesgebiet unerläßlich ist, nicht für einen Mobilfunkanbieter entscheiden, der nur eine eingeschränkte Flächendekkung bietet. Die Präferenzstruktur dieses Kunden ist dadurch gekennzeichnet, daß

---

[1] Vgl. Brockhoff (1999), S. 13.

die Flächendeckung des Netzanbieters ein kaufentscheidendes Kriterium darstellt. In diesem Beispiel wird deutlich, daß die Präferenzstrukturen der Konsumenten nicht wie die Kaufentscheidung nur Auskunft darüber geben, *ob* ein Produkt letztendlich gekauft wird, sondern sie geben Antworten auf die Frage, *warum* ein Produkt gekauft wird oder eben nicht. Die Aufgabe der Marktforschung ist daher, Methoden bereitzustellen, anhand derer die Präferenzstrukturen der Nachfrager erfaßt und abgebildet werden können.

Eine der bedeutendsten Methoden zur Messung von Nachfragerpräferenzen ist die **Conjoint-Analyse**. Sie zählt zweifelsohne zu den populärsten multivariaten Analysemethoden im Marketing. Dies gilt sowohl aus Sicht der Wissenschaft als auch für die kommerzielle Anwendung des Verfahrens in der Marketing-Praxis. Die Conjoint-Analyse ist mittlerweile fester Bestandteil in jedem Lehrbuch der Marktforschung und ist in zunehmendem Maße Forschungsgegenstand zahlreicher Veröffentlichungen in wissenschaftlichen Zeitschriften.[2] Gleichzeitig hat kaum ein anderes Analyseinstrument, das in der wissenschaftlichen Forschung entwickelt wurde, eine so weite Akzeptanz und Verbreitung in der Praxis gefunden. Bereits Mitte der achtziger Jahre schätzen Wittink und Cattin (1989) die Zahl der kommerziellen Anwendungen allein in den USA auf 400 im Jahr.[3] Für Europa konnten in dem Zeitraum von 1986 bis 1991 nahezu 1000 von Marktforschungsunternehmen durchgeführte Conjoint-Studien nachgewiesen werden.[4] Mittlerweile dürfte diese Zahl noch um ein Vielfaches höher liegen.[5]

Grundidee der traditionellen Conjoint-Analyse ist es, die Präferenzstruktur von Konsumenten auf indirektem Weg zu ermitteln. Die Konsumenten werden nach Präferenzurteilen für ganzheitlich beschriebene Produktalternativen gefragt. Es ist die Aufgabe des Marktforschers, auf Basis der erhobenen Präferenzen den Nutzenbeitrag jeder einzelnen Produkteigenschaft zum Gesamturteil zu ermitteln.[6] Diese Zerlegung der erhobenen Gesamturteile in einzelne Teilkomponenten

---

2 Vgl. Voeth (1999), S. 154.
3 Vgl. Wittink/Cattin (1989), S. 92.
4 Vgl. Wittink/Vriens/Burhenne (1994), S. 43.
5 Vgl. Voeth/Hahn (1998), S. 119 f.
6 Vgl. Green/Srinivasan (1978), S. 104.

wird auch als dekompositionelles[7] Vorgehen bezeichnet und stellt das wesentliche Merkmal der Conjoint-Analyse dar. Immer häufiger, vor allem bei Relevanz vieler Produkteigenschaften, stößt die traditionelle Conjoint-Analyse allerdings an ihre Grenzen. Die ganzheitliche Beschreibung der Produktkonzepte - entsprechend der realen Kaufentscheidung - erhöht zwar den Realitätsbezug der Aufgabenstellung, kann aber insbesondere bei mehr als 6 Produkteigenschaften aufgrund einer kognitiven Überlastung der befragten Personen zu Validitätsproblemen führen.[8] Da in praktischen Anwendungen häufig eine größere Anzahl an Produkteigenschaften untersucht werden soll, sind vor allem für solche Anwendungen hybride Verfahren entwickelt worden.[9] **Hybride Verfahren** kombinieren einen kompositionellen und einen dekompositionellen Befragungsschritt miteinander. Die Auskunftspersonen geben hierbei im kompositionellen Teil, auch **Self-Explicated-Methode** genannt, ein direktes Urteil über die Wichtigkeit der Eigenschaften und deren Ausprägungen ab.[10] In einem zweiten, dekompositionellen Schritt werden ausgewählte Eigenschaftskombinationen wie bei einer traditionellen Conjoint-Analyse bewertet. Die Einfachheit der Self-Explicated-Methode wird mit der realistischeren Aufgabe der Conjoint-Analyse kombiniert. Ziel ist es, die Vorteile beider Verfahren zu nutzen und die Komplexität der Aufgabenstellung für die Probanden zu reduzieren.[11] Abbildung 1-1 beschreibt den Zusammenhang der beschriebenen Ansätze der Präferenzmessung.

---

7 Der Begriff „dekompositionell" wurde von dem englischen Verb „to decompose" (deutsch: „zerlegen") abgeleitet.

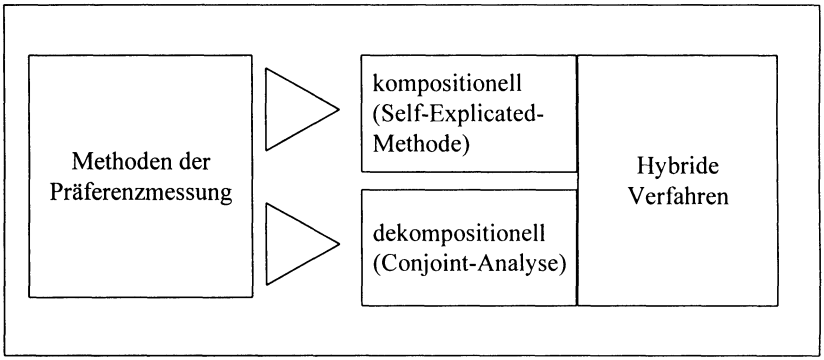
8 Green/Srinivasan (1990), S. 8 f.

9 Vgl. Green (1984), S. 156.

10 Die Self-Explicated-Methode geht als kompositionelles Verfahren genau umgekehrt vor wie die Conjoint-Analyse. Statt globaler Gesamturteile bezüglich mehrerer Produktalternativen zu erheben, werden die Probanden direkt nach der Wichtigkeit der einzelnen Eigenschaften und deren Ausprägungen gefragt. Nachträglich werden diese Einzelurteile für die betrachteten Eigenschaften vom Marktforscher zu einem Gesamturteil für alternative Produkte kombiniert (vgl. Hoepfl/Huber (1970), S. 408; Srinivasan (1988), S. 296).

11 Vgl. Green (1984), S. 156.

Abbildung 1-1: Ansätze zur Präferenzmessung



Quelle: Eigene Erstellung

Mit der Entwicklung hybrider Verfahren kommt auch der computergestützten Datenerhebung eine immer größere Bedeutung zu. Computergestützte Interviews können, insbesondere durch Anpassung des Interviewablaufs an vorangegangene Antworten, die Effizienz der Datenerhebung erhöhen.[12] Dies zeigt sich auch daran, daß die mit Abstand am weitesten verbreitete Form der hybriden Conjoint-Analyse die von Johnson (1987) in Verbindung mit Sawtooth Software entwickelte **Adaptive Conjoint Analysis** (im folgenden: **ACA**) ist. Sie stellt die bisher einzige kommerzielle Conjoint-Software dar, die sich durch einen adaptiven Interviewablauf auszeichnet.[13] Der Präferenzbildungsprozeß wird dabei auf Basis individuell wichtiger Merkmale analysiert und bietet damit vor allem den Vorteil, Studien mit einer großen Anzahl von Eigenschaften durchführen zu können.[14] Nicht zuletzt die komfortable Integration von Datenerhebung, Auswertung und Marktsimulation in einem Softwarepaket begründet die (zunehmende) Attraktivität der ACA.[15] Bereits 1994 kam sie in 42% aller in Europa durchgeführten Conjoint-Studien zum Einsatz.[16]

Der Interviewablauf der ACA umfaßt 5 aufeinander aufbauende Datenerhebungsschritte.[17] Im Self-Explicated-Teil (Schritt 1 bis 3) werden die Probanden direkt

---

12 Vgl. Bond (1991), S. 749; Müller/Kesselmann (1996), S. 193.
13 Vgl. Carmone/Schaffer (1995), S. 117.
14 Vgl. Johnson (1987), S. 259.
15 Vgl. Carmone (1987), S. 327 und Green/Srinivasan (1990), S. 11.
16 Vgl. Wittink/Vriens/Burhenne (1994), S. 45.
17 Zum Ablauf der ACA vgl. Johnson (1987), S. 259 ff.

nach Bewertungen für alle Eigenschaftsausprägungen und nach den Wichtigkeiten der einzelnen Eigenschaften gefragt. Der sich anschließende dekompositionelle Interviewteil (Schritt 4) besteht aus einer wiederholten Paarvergleichsaufgabe. Den Probanden werden jeweils zwei Produktalternativen präsentiert, die anhand von zwei bis fünf Eigenschaften beschrieben sind. Für jeden Paarvergleich soll auf einer Rating-Skala angegeben werden, welches der beiden Produkte bevorzugt wird. Die Besonderheit der ACA besteht darin, daß die Konzeption der Paarvergleiche individuell an vorangegangene Antworten angepaßt und damit für jede Testperson individuell gestaltet wird. Das Interview endet jeweils mit der Bewertung ganzheitlich beschriebener Produktkonzepte, für die eine Kaufwahrscheinlichkeit von 0% bis 100% anzugeben ist (Schritt 5). Zur Bestimmung der Präferenzstrukturen führt die ACA jeweils zwei getrennte Schätzungen durch: Sie resultieren zum einen aus dem Self-Explicated-Teil und zum anderen aus den Ergebnissen der Paarvergleichsaufgabe. Der fünfte Schritt dient der optimalen Gewichtung dieser beiden Schätzergebnisse.

Trotz ihrer weiten Verbreitung weist die ACA verschiedene Schwachstellen in der Vorgehensweise auf. Green/Krieger/Agarwal (1991) kritisieren hauptsächlich die wiederholte Verwendung von Paarvergleichen. Sie bewerten diese als sehr unrealistisch und bemängeln, daß die Paarvergleiche aufgrund ihrer Konstruktion zum Teil sehr schwer zu bewältigen sind.[18] Um die Effizienz der Datenerhebung zu erhöhen, wird die Paarvergleichsaufgabe immer so konstruiert, daß die beiden Produkte in ihrer Attraktivität so ähnlich wie möglich sind.[19] Die Schwierigkeit für die Auskunftspersonen besteht allerdings darin, daß sie sich wiederholt zwischen Alternativen entscheiden müssen, zwischen denen sie eigentlich nahezu indifferent sind.[20] Louviere/Gaeth (1988) betonen zudem in einer allgemeinen Diskussion um den Einsatz von Paarvergleichen, daß die Zahl erforderlicher Vergleiche in Abhängigkeit der Anzahl zu beurteilender Alternativen sehr schnell ansteigt.[21] Im Zusammenhang mit der Paarvergleichsaufgabe weisen Srinivasan/Park (1997) gleichzeitig darauf hin, daß diese nicht ausreichend viele Daten-

---

18 Vgl. Green/Krieger/Agarwal (1991), S. 220.
19 Vgl. Johnson (1987), S. 262.
20 Vgl. Green/Krieger/Agarwal (1991), S. 216.
21 Vgl. Louviere/Gaeth (1988), S. 63.

punkte für eine verzerrungsfreie Schätzung liefert.[22] Als letzter Kritikpunkt sei angemerkt, daß die Anwenderfreundlichkeit der ACA durch Integration von Datenerhebung und deren Auswertung gleichzeitig die Gestaltungsmöglichkeiten enorm einschränkt. So besteht weder eine Einflußmöglichkeit auf die Konstruktion der Paarvergleiche, die den Probanden präsentiert werden, noch auf den Schätzprozeß, auf dem die Auswertung der Präferenzdaten beruht.[23] Über diese theoretischen Kritikpunkte an der ACA hinaus rechtfertigen auch bisherige empirische Ergebnisse die Dominanz des Verfahrens kaum. Agarwal (1989) berichtet, daß durch die zusätzlichen Informationen aus der Paarvergleichsaufgabe keine signifikante Verbesserung hinsichtlich der Prognoseergebnisse erzielt werden konnte.[24] Auch bei zahlreichen empirischen Validitätsvergleichen ist die ACA einer oder mehreren der untersuchten alternativen Methoden zur Präferenzmessung unterlegen.[25]

Faßt man die einführende Diskussion zusammen, können drei wesentliche Punkte festgehalten werden:

- Kenntnisse über die Nachfragerpräferenzen sind von zentraler Bedeutung für die Produkt-Manager, um Produkte erfolgreich am Markt einzuführen, zu positionieren und letztendlich zu etablieren.
- Der Einsatz der traditionellen Conjoint-Analyse als Instrument der Marktforschung zur Messung von Nachfragerpräferenzen ist insbesondere bei Anwendungen mit vielen Produkteigenschaften problematisch. Zur Lösung dieser Probleme sind neue, vor allem hybride Verfahren entwickelt worden.
- Besonderer Beliebtheit erfreut sich die Adaptive Conjoint Analysis (ACA) als bisher einzige computergestützte hybride Methode zur Präferenzmessung. Trotz methodischer Schwächen und bisher kaum überzeugenden empirischen Ergebnissen hat sie sich nahezu als Standardinstrument zur Präferenzmessung etabliert.

---

22 Vgl. Srinivasan/Park (1997), S. 286.

23 Vgl. Carmone (1987), S. 327.

24 Vgl. Agarwal (1989), S. 8.

25 Vgl. hierzu die Ergebnisse der Studien von Finkbeiner und Platz (1986), Agarwal (1988) (beide zitiert nach Green/Srinivasan (1990), S. 11), Green, Krieger und Agarwal (1993), S. 220 und Pullmann, Dodson und Moore (1999), S. 136.

Die Akzeptanz der Conjoint-Analyse und die zunehmende Vielfalt an Einsatzgebieten der Präferenzmessung stellen gleichzeitig immer höhere Anforderungen an die Methoden.[26] Für die praktische Anwendung müssen sie vor allem für Fragestellungen mit vielen Produkteigenschaften geeignet sein, wodurch computergestützten Verfahren eine immer größere Bedeutung zukommt. Die weite Verbreitung der Adaptive Conjoint Analysis (ACA) verdeutlicht diese Entwicklung. Aufgrund der aufgezeigten Schwächen der ACA stellt sich jedoch die Frage, ob alternative Verfahren besser geeignet sind, die Präferenzstrukturen der Nachfrager zu messen.

1.2 Ziele der Arbeit

Im Rahmen dieser Arbeit sollen neue Methoden zur Präferenzmessung entwickelt werden. Mit dem klaren Ziel der praktischen Anwendung müssen diese auch bei Studien mit vielen Produkteigenschaften einsetzbar sein. Aus diesem Grund werden computergestützte hybride Methoden entwickelt, die ebenfalls einen individualisierten Interviewablauf ermöglichen. Mit dieser Individualisierung wird eine Verringerung der Aufgabenkomplexität sowie des Erhebungsaufwands erreicht, da im dekompositionellen Befragungsteil sowohl weniger als auch für jeden Probanden die tatsächlich kaufrelevanten Eigenschaften und Eigenschaftsausprägungen berücksichtigt werden. Um die Schwachstellen der ACA zu überwinden, werden insbesondere im Conjoint-Teil statt einer wiederholten Paarvergleichsaufgabe alternative Vorgehensweisen in der Datenerhebung gewählt. Im Rahmen einer empirischen Anwendung der neuen Methoden und der ACA soll getestet werden, ob die neuen Verfahren die Schwachstellen der ACA überwinden können. Somit verfolgt die vorliegende Arbeit zwei wesentliche Ziele:

- ***Entwicklung neuer computergestützter hybrider Conjoint-Analysen***

Entwickelt werden zwei neue computergestützte Conjoint-Analysen. Aufbauend auf einem von Srinivasan und Park (1997) vorgeschlagenen Ansatz, der **Customized Conjoint Analysis** (im folgenden **CCA**), wird die **Customized Computerized Conjoint Analysis** (im folgenden: **CCC**) entwickelt. Entsprechend dem Ansatz von Srinivasan und Park (1997) besteht der dekompositionelle Befragungsteil aus einer Rating- oder Ranking-Aufgabe aller zu bewertenden Produk-

---

[26] Zu möglichen Einsatzgebieten der Conjoint-Analyse vgl. Gustafsson/Herrmann/Huber (2000), S. 6 f.

te.[27] Im Vergleich dazu werden bei der ACA jeweils nur zwei Produktbeschreibungen simultan bewertet. Die Produkte werden dabei anhand der individuell wichtigsten Eigenschaften beschrieben, die im Rahmen der vorangegangenen kompositionellen Befragung identifiziert werden. Mit der Implementierung eines computergestützten Interviews wird ein wesentlicher Nachteil der CCA, bedingt durch eine Zeitverzögerung von zwei Wochen zwischen kompositionellem und dekompositionellem Befragungsteil, überwunden.[28]

Mit der **Choice-Oriented Individualized Conjoint Analysis** (im folgenden: **CHIC**) wird eine zweite computergestützte hybride Conjoint-Variante vorgeschlagen. Die Gestaltung des dekompositionellen Befragungsteils ist dabei an die **Choice-Based Conjoint Analysis** (in folgenden **CBC**) angelehnt, die sich zunehmender Beliebtheit im Marketing erfreut.[29] Die Besonderheit der CBC besteht darin, daß die Präferenzen der Probanden in Form einer Auswahlentscheidung erfragt werden.[30] Präsentiert werden hierfür verschiedene Teilmengen an Produktkonzepten, aus denen sich die Probanden für jeweils eine Alternative entscheiden können.[31] In Anlehnung daran werden im Conjoint-Teil der CHIC den Probanden sogenannte Choice-Sets zur Bewertung präsentiert. Diese Choice-Sets bestehen immer aus einer Teilmenge von Produkten, die wie bei der CCC anhand der individuell wichtigsten Eigenschaften beschrieben werden. Damit sind bei der CHIC im Gegensatz zur CCC weniger Produkte simultan zu bewerten, gleichwohl aber mehr als bei der ACA.

- ***Empirische Validitätsprüfung der entwickelten Verfahren***

Über die Entwicklung neuer Methoden hinaus muß deren Anwendbarkeit in der Praxis gewährleistet sein. Ein zweites Ziel der vorliegenden Arbeit stellt daher eine empirische Überprüfung der neu entwickelten Methoden dar, wobei gleichzeitig ein Vergleich mit bereits etablierten Verfahren vorgenommen wird. So wird in die empirische Studie das Standardverfahren ACA integriert, das als eine Referenzgröße für die Ergebnisgüte dienen soll. Darüber hinaus werden die com-

---

27 Zur Grundidee der CCA vgl. Srinivasan/Park (1997), S. 286.
28 Vgl. Srinivasan/Park (1997), S. 287.
29 Vgl. hierzu Green/Srinivasan (1990); Schubert (1995) und Weiber/Rosendahl (1997).
30 Vgl. Weiber/Rosendahl (1997), S. 109.
31 Vgl. Schubert (1995), S. 381.

putergestützten hybriden Methoden mit einfacheren kompositionellen Self-Explicated-Modellen verglichen. Zwar wird der Conjoint-Analyse insbesondere durch ihren stärkeren Realitätsbezug gegenüber der Self-Explicated-Methode aus theoretischer Sicht stets eine höhere Prognosevalidität zugestanden,[32] jedoch kann diese Überlegenheit bisher empirisch nicht eindeutig nachgewiesen werden. Sattler und Hensel-Börner (2000) zeigen auf Basis eines Vergleichs bisheriger Befunde, daß nur in wenigen Fällen die Conjoint-Analyse im Vergleich zu Self-Explicated-Methoden signifikant validere Ergebnisse aufweist. Vielmehr zeigen sich in fast 80% der betrachteten Befunde keine signifikanten Unterschiede oder sogar eine Überlegenheit der Self-Explicated-Modelle.[33] Vor diesem Hintergrund wird die Self-Explicated-Methode neben der ACA als zweite Referenzgröße für die Ergebnisgüte der neu entwickelten Verfahren herangezogen.

1.3 Aufbau der Arbeit

Im Anschluß an diese Einleitung werden in Kapitel 2 die verschiedenen Ansätze zur Präferenzmessung dargestellt. Für alle drei Verfahrensbereiche kompositionell, dekompositionell und hybrid (vgl. Abbildung 1-1), werden die unterschiedlichen Varianten systematisiert und deren Besonderheiten abgeleitet. Zunächst werden dabei die beiden gegensätzlichen Vorgehensweisen: kompositionell (die Self-Explicated-Methode) und dekompositionell (die Conjoint-Analyse) separat betrachtet, wobei neben der Darstellung der verschiedenen Varianten auch die Ergebnisse bisheriger empirischer Vergleichsstudien diskutiert werden. Eine Gegenüberstellung der beiden Vorgehensweisen wird anschließend sowohl auf Basis theoretischer Vor- und Nachteile der Verfahren als auch durch Auswertung empirischer Befunde vorgenommen. Als dritter Bereich werden die hybriden Methoden als Kombination aus Self-Explicated-Methode und Conjoint-Analyse vorgestellt. Über die Systematisierung der verschiedenen Formen wird auch hier das empirische Abschneiden der Verfahren, vor allem im Vergleich zu den beiden anderen Ansätzen, rein kompositionell und rein dekompositionell, zusammengefaßt.

---

32 Vgl. Oppedijk van Veen/Beazly (1977), S. 4; Green/Goldberg/Montemayer (1981), S. 33; Green/Srinivasan (1990), S. 13; Sattler/Hensel-Börner (2000), S. 124.

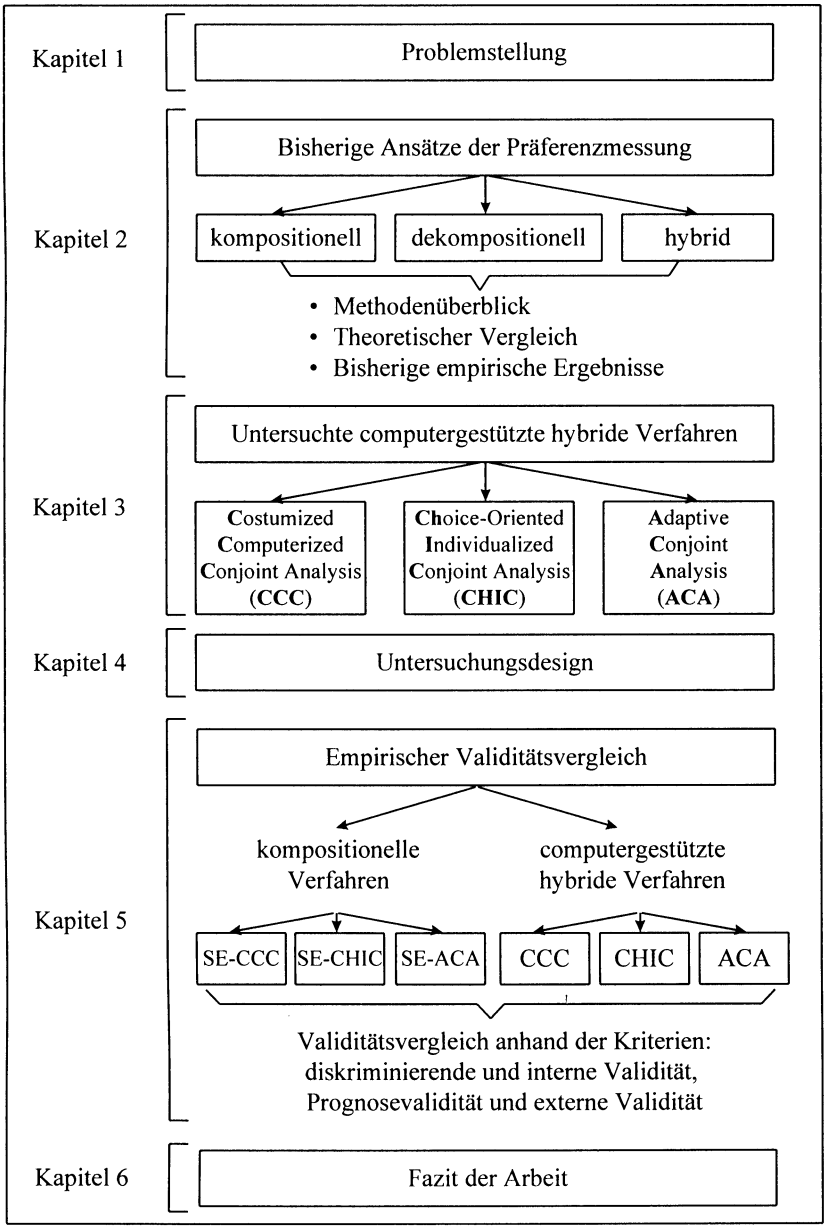
33 Vgl. Sattler/Hensel-Börner (2000), S. 131.

Im dritten Kapitel werden die drei computergestützten hybriden Verfahren CCC, CHIC und ACA ausführlich dargestellt. Schwerpunkte der Darstellung sind jeweils die Vorgehensweise bei der Datenerhebung und das Schätzverfahren zur Ermittlung der Präferenzstrukturen. Alle drei Verfahren werden in einer empirischen Anwendung zum einen miteinander und zum anderen mit der Self-Explicated-Methode verglichen. Das Untersuchungsdesign dieser Studie ist Gegenstand des vierten Kapitels.

Für die vergleichende Validitätsprüfung der verschiedenen Verfahrensvarianten werden unterschiedliche Beurteilungskriterien herangezogen. Neben der diskriminierenden Validität wird die interne Validität als Indikator für die jeweilige Modellgüte analysiert. Im Mittelpunkt des Verfahrensvergleichs steht die Prüfung der Prognosegüte. Auf Basis der erhobenen Präferenzurteile werden für die in der Untersuchung betrachteten Marken Marktanteile prognostiziert. Zur Überprüfung der Prognosevalidität dient der Vergleich mit den in einer simulierten Kaufentscheidung (sog. Holdout-Sample) beobachteten Marktanteilen. Durch einen Vergleich mit realen Marktanteilsdaten wird zudem die externe Validität der verschiedenen Verfahren getestet. Die Vorgehensweise bei der Validierung der Verfahren und die jeweiligen Ergebnisse der Validitätsprüfungen werden in Kapitel 5 dargestellt. Das fünfte Kapitel ist dabei gemäß der eingesetzten Gütemaße untergliedert. In Abschnitt 5.5 werden die Ergebnisse hinsichtlich aller eingesetzten Gütemaße noch einmal zusammengefaßt.

Die Arbeit endet im sechsten Abschnitt mit einem Fazit. Abbildung 1-2 gibt einen Überblick über den Aufbau der Arbeit.

Abbildung 1-2: Aufbau der Arbeit



Quelle: Eigene Erstellung

2 Ansätze zur Messung von Nachfragerpräferenzen

Im folgenden wird ein Überblick über bestehende Ansätze zur Präferenzmessung gegeben, bei denen die Präferenzen von Probanden in Form von Befragung erhoben werden.[34] Schwerpunkt der Ausführungen ist die Vorgehensweise bei der Datenerhebung. Unter dem Begriff Datenerhebung wird im Rahmen dieser Arbeit sowohl die Befragungsform (direkt oder indirekt) als auch die konkrete Gestaltung der Präferenzabfrage verstanden. Über die Darstellung der verschiedenen Ansätze und einer Diskussion der jeweiligen Vor- und Nachteile hinaus, werden bisherige empirische Forschungsergebnisse zusammengefaßt. Berücksichtigt werden jeweils die Studien, bei denen vornehmlich ein Vergleich der Datenerhebung Untersuchungsgegenstand war oder im Rahmen anderer Untersuchungsschwerpunkte implizit getestet wurde. Um insgesamt einen breiten Überblick über empirische Forschungsergebnisse zu den verschiedenen Ansätzen der Präferenzmessung geben zu können, werden alle seit 1980 veröffentlichten Studien in den internationalen Zeitschriften „Journal of Marketing Research", „Journal of Marketing", „International Journal of Research in Marketing", „Marketing Letters" und „Marketing Science" herangezogen. Für die deutschsprachige Literatur werden die Untersuchungen betrachtet, die in der „Zeitschrift für Betriebswirtschaft", „Zeitschrift für betriebswirtschaftliche Forschung" , „Marketing Zeitschrift für Forschung & Praxis" und „Die Betriebswirtschaft (DBW)" veröffentlicht sind. Zusätzlich werden häufig zitierte Studien und Arbeitspapiere ausgewertet.

Der Aufbau dieses Kapitels ist an die Dreiteilung der Ansätze entsprechend Abbildung 1-1 angelehnt. Zunächst werden die kompositionellen (Abschnitt 2.1) und die dekompositionellen Ansätze (Abschnitt 2.2) als gegensätzliche Vorgehensweisen separat betrachtet. Der Darstellung der einzelnen Varianten innerhalb dieser beiden Bereiche folgt eine Auswertung empirischer Befunde bisheriger Vergleichsstudien. In Abschnitt 2.3 folgt der getrennten Betrachtung eine ver-

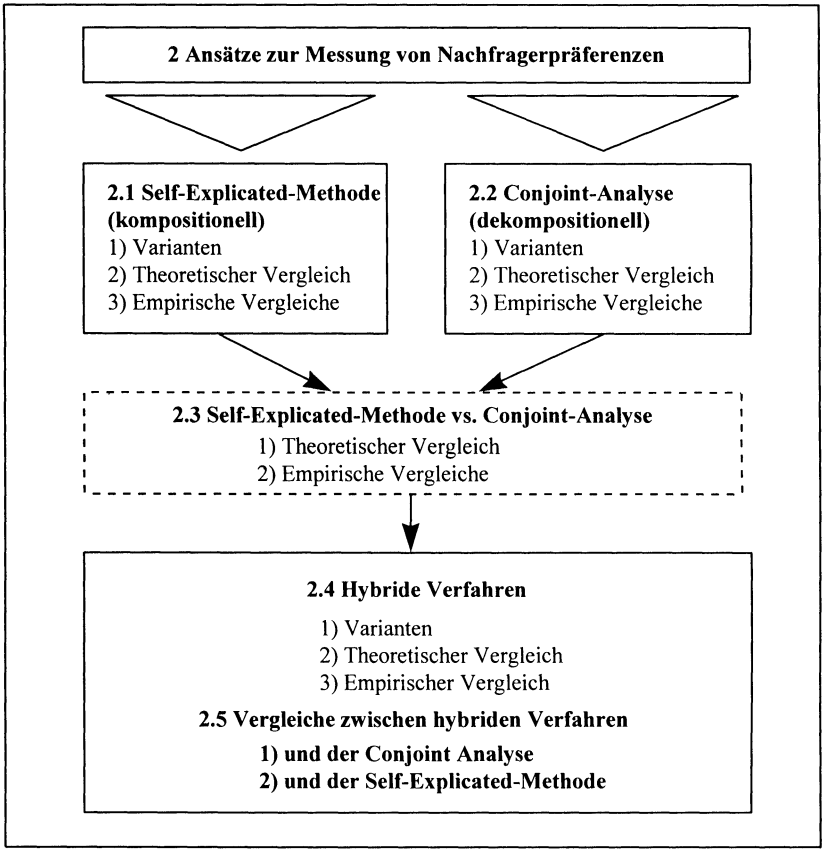
---

[34] Eine weitere Möglichkeit zur Präferenzmessung stellt die Beobachtung und Auswertung tatsächlicher Marktdaten dar, welche in der vorliegenden Arbeit nicht weiter betrachtet werden soll.

gleichende Diskussion der beiden Bereiche, die in einen theoretischen und empirischen Vergleich unterteilt ist.

Als Kombination der beiden Vorgehensweisen - kompositionell und dekompositionell - bilden die hybriden Verfahren einen dritten großen Verfahrensbereich der Ansätze zur Präferenzmessung. Die verschiedenen Varianten hybrider Verfahren werden ebenfalls dargestellt (2.4) und abschließend mit den beiden anderen Bereichen verglichen (Abschnitt 2.5). Die folgende Abbildung 2-1 gibt einen Überblick über den Aufbau dieses Kapitels.

Abbildung 2-1: Überblick über den Aufbau von Kapitel 2



Quelle: Eigene Erstellung

2.1 Kompositionelle Verfahren (Self-Explicated-Methode)

Kompositionelle Methoden zur Präferenzmessung, allgemein auch Self-Explicated-Methoden genannt, sind maßgeblich dadurch gekennzeichnet, daß die Präferenzstruktur von Konsumenten direkt erhoben wird. Dabei werden Bewertungen einzelner Eigenschaften und deren Ausprägungen explizit von den Probanden erfragt und anschließend zu einem Gesamturteil zusammengefaßt.[35] Grundlage hierfür ist das multiattributive Modell, bei dem ein Produkt als Bündel von Eigenschaften charakterisiert wird.[36] Das lineare multiattributive Präferenzmodell[37] kann formal wie folgt dargestellt werden:[38]

$$(2\text{-}1) \qquad U_{im} = \sum_{j \in J} \sum_{k \in K_j} \alpha_{ijk} \cdot x_{jkm} \qquad (i \in I;\ m \in M)$$

wobei:

U_{im}: Geschätzter Gesamtnutzen für die m-te Alternative für den i-ten Probanden,

α_{ijk}: Bewertung für die k-te Ausprägung der j-ten Eigenschaft durch den i-ten Probanden,

x_{jkm}: k-te Ausprägung der j-ten Eigenschaft bei der m-ten Alternative,

J: Indexmenge der Eigenschaften,

K_j: Indexmenge der Eigenschaftsausprägungen der j-ten Eigenschaft,

I: Indexmenge der Probanden,

M: Indexmenge der Alternativen.

Dieses additive Nutzenmodell basiert auf der Annahme, daß sich der Gesamtnutzen einer Alternative als Summe der Teilnutzenwerte für die Eigenschaftsausprägungen, die diese Alternative beschreiben, ergibt. Die Teilnutzenwerte stellen da-

---

35 Vgl. Jain/Mahajan/Malhotra (1979), S. 248.

36 Vgl. Wilkie/Pessemier (1973), S. 428.

37 Das multiattributive Modell geht ursprünglich auf die Einstellungsforschung zurück. In Anlehnung daran wird auch von der Einstellung für ein Produkt gesprochen (vgl. z.B. Wilkie/Pessemier (1973)). In der Marketing-Literatur hat sich der Begriff der Präferenz, gemessen durch den Nutzen, durchgesetzt und wird im folgenden angewendet.

38 Die formale Darstellung ist angelehnt an Jain/Mahajan/Malhotra (1979), S. 248.

bei numerische Werte dar, anhand derer die subjektiven Bewertungen der Eigenschaftsausprägungen abgebildet werden. Sie können für jeden Probanden berechnet werden und aufgrund eines einheitlichen Skalenniveaus für alle Eigenschaftsausprägungen direkt miteinander verglichen werden.[39] Alternativ besteht auch die Möglichkeit, die einzelnen Teilnutzenwerte unter Annahme eines nicht-linearen Modells zu kombinieren. Nicht-lineare Modelle zeichnen sich insbesondere dadurch aus, daß neben den Haupteffekten zusätzlich auch Wechselwirkungen zwischen den Eigenschaften berücksichtigt werden können.[40]

Neben diesem über alle kompositionellen Methoden konsistenten Grundgedanken der direkten Befragungsform, unterscheiden sich die verschiedenen Verfahrensvarianten insbesondere im Umfang der erhobenen Präferenzurteile. Bei der einfachsten Form kompositioneller Verfahren wird lediglich die Wichtigkeit der zu untersuchenden Eigenschaften erhoben, wobei die jeweiligen Ausprägungen der Eigenschaften ignoriert werden.[41] Diese **traditionelle Form**[42] kompositioneller Verfahren findet sich auch in anderen Forschungsfeldern. In der Entscheidungstheorie wird beispielsweise das sog. Direct-Ratio-Verfahren[43] eingesetzt, um die Wertfunktion eines Entscheiders zu ermitteln.[44] Die mit dieser extrem einfachen Methode verbundenen Probleme, insbesondere der Bandbreiteneffekt[45], haben

---

39 Zur Definition der Teilnutzenwerte vgl. Thomas (1979), S. 199.

40 Eine sehr gute Gegenüberstellung linearer und nicht-linearer Präferenzmodelle findet sich bei Reiners (1996), S. 58 f.

41 Für die formale Darstellung impliziert dies, daß in Formel (1) der Index k für die Ausprägung der j-ten Eigenschaft entfällt. Vgl. hierzu auch die Darstellung bei Wilkie/Pessemier (1973), S. 429.

42 Jain/Mahajan/Malhotra (1979), S. 248 sprechen bei dieser vereinfachten Form von dem traditionellen kompositionellen Modell.

43 Das Direct-Ratio-Verfahren entspricht der vorgestellten traditionellen Form der Self-Explicated-Methode.

44 Vgl. Eisenführ/Weber (1994), S. 128 f.

45 Das Prinzip der Bandbreitensensitivität besagt, daß die Bedeutung einer Eigenschaft von den Bandbreiten der jeweiligen Eigenschaftsausprägungen abhängt (vgl. Fischer (1995), S. 252). Dementsprechend muß sich eine Änderung der Bandbreite der Eigenschaftsausprägungen im entsprechenden Verhältnis auf das Gewicht dieser Eigenschaft auswirken. Ein Bandbreiteneffekt liegt dann vor, wenn die tatsächliche Variation des Bedeutungsgewichtes stark von der theoretisch geforderten Variation abweicht. In diesem Fall ist der Entscheider nicht in der Lage, die Bedeutungsgewichte konsistent zu ermitteln (vgl. von Nitzsch/Weber (1991), S. 972). Da bei dieser traditionellen Variante der kompositionellen Modelle die Wichtigkeit unabhängig von den Eigenschaftsausprägungen abgefragt wird, ist die Problematik der Bandbreitensensitivität von zentraler Bedeutung. Zur Bandbreitensensitivität der verschiedenen Verfahren zur Präferenzmessung vgl. Sattler/Gedenk/Hensel-Börner (2000).

dazu geführt, daß häufig eingesetzte kompositionelle Verfahren sowohl eine Bewertung der Wichtigkeit der Eigenschaften als auch eine Beurteilung der Eigenschaftsausprägungen umfassen. Typischerweise werden zunächst Beurteilungen für die Ausprägungen einer jeden Eigenschaft abgefragt. In einem zweiten Schritt geben die Probanden in Abhängigkeit der von ihnen bereits beurteilten Ausprägungen die Wichtigkeit für die einzelnen Eigenschaften an. Eine solche Vorgehensweise wird im allgemeinen als **gewichtete Self-Explicated-Methode** bezeichnet, wobei die konkrete Ausgestaltung dieser beiden Teilschritte variieren kann.[46] Während im Self-Explicated-Teil der Adaptive Conjoint Analysis (ACA) die Bedeutung jeder Eigenschaft separat erhoben wird[47], schlägt Srinivasan (1988) vor, bei der Wichtigkeit der Eigenschaften zunächst eine sog. "kritische" Eigenschaft zu identifizieren. Diese "kritische" Eigenschaft soll die für den Probanden wichtigste Eigenschaft sein und bei der Beurteilung aller weiteren als "Anker" dienen. Ziel dieser Modifikation ist, die Wahrscheinlichkeit zu erhöhen, die Bedeutung der verschiedenen Eigenschaften zueinander tatsächlich auf einer Ratio-Skala zu ermitteln.[48] Die metrische Interpretation der daraus ermittelten Teilnutzenwerte setzt ein solches Skalenniveau voraus.

Wie bereits erläutert, stellt die Abfrage der Wichtigkeiten der Eigenschaften, unabhängig von einer Einschätzung der Ausprägungen, eine vereinfachte Form der Self-Explicated-Methode dar und entspricht der traditionellen Vorgehensweise. Umgekehrt wird bei einer alternativen Vereinfachung explizit auf diesen Schritt der Befragung verzichtet. Von den Probanden werden lediglich die verschiedenen Eigenschaftsausprägungen beurteilt. Unter Annahme gleicher Gewichte für alle Eigenschaften werden anschließend die Teilnutzenwerte ermittelt.[49]

Neben diesen Vereinfachungen kann die Self-Explicated-Methode auch um eine vorherige Abfrage völlig inakzeptabler Eigenschaftsausprägungen erweitert werden. Diese Form von Self-Explicated-Verfahren wird in der Literatur unter dem Begriff **Conjunctive-Compensatory Self-Explicated-Methode** eingeführt.[50]

---

46 Vgl. Green/Srinivasan (1990), S. 9.

47 Zum Ablauf der Adaptive Conjoint Analysis vgl. Kapitel 3.3 dieser Arbeit.

48 Vgl. Srinivasan (1988), S. 296.

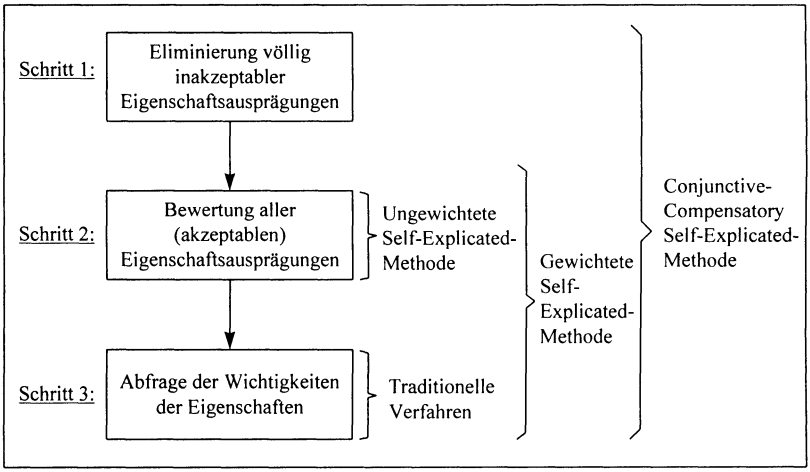
49 Vgl. Green/Krieger/Agarwal (1993), S. 373.

50 Zur Conjunctive-Compensatory Self-Explicated-Methode vgl. Srinivasan (1988).

Eine solche Zweistufigkeit basiert auf der Überlegung, daß nicht akzeptable Eigenschaftsausprägungen für Konsumenten k.o.-Kriterien darstellen können und demzufolge gar kein tatsächlicher Trade-Off mehr zwischen den verschiedenen Alternativen vorgenommen wird. Sofern solche absolut inakzeptablen Ausprägungen eliminiert werden, kann mit der zweiten "compensatory" Phase eher das tatsächliche Wahl- bzw. Entscheidungsverhalten von Nachfragern abgebildet werden.[51]

Abbildung 2-2 gibt einen zusammenfassenden Überblick über die verschiedenen Varianten kompositioneller Ansätze, differenziert nach den enthaltenen Ablaufschritten bei der Datenerhebung.

Abbildung 2-2: Ablaufschritte kompositioneller Verfahren zur Präferenzmessung



Quelle: Eigene Erstellung

Die Ergebnisse bisheriger **empirischer Validitätsvergleiche** der verschiedenen Varianten der Self-Explicated-Methode sind widersprüchlich. Während Agarwal/Green (1991) noch eine signifikante Überlegenheit hinsichtlich der Prognosevalidität des gewichteten (Schritt 2 und 3 in Abbildung 2-2) gegenüber dem ungewichteten Modell (nur Schritt 2 gemäß Abbildung 2-2) nachweisen kön-

[51] Vgl. Dorsch/Teas (1992), S. 38 f.

nen,[52] variieren die Ergebnisse je nach Operationalisierung der Beurteilungskriterien bei Green/Krieger/Agarwal (1993) und lassen damit keine abschließende Aussage zugunsten einer der beiden Methoden zu.[53]

Entsprechendes gilt für die Erweiterung der Befragung um den ersten Befragungsschritt, sprich der Eliminierung völlig inakzeptabler Eigenschaftsausprägungen (vgl. Abbildung 2-2). So testet Srinivasan (1988) die von ihm vorgeschlagene Conjunctive-Compensatory Self-Explicated-Methode (im folgenden: CC-SE) explizit gegenüber der üblichen Self-Explicated-Variante (ohne Eliminierung inakzeptabler Ausprägungen) und kann für die CC-SE eine (schwach) signifikant ($p<0,1$) höhere Prognosevalidität (gemessen durch externe Validität) nachweisen.[54] Allerdings spricht gegen die Eliminierung inakzeptabler Eigenschaftsausprägungen jedoch die Gefahr, daß Probanden häufig Ausprägungen zu schnell als solche einstufen. Obwohl Produktalternativen z.T. durch inakzeptabel eingestufte Ausprägungen definiert sind, werden sie bei einer anschließenden ganzheitlichen Beurteilung gar nicht absolut abgelehnt. Empirische Belege für solch ein inkonsistentes Verhalten liefern u.a. Krieger (1986)[55], Green/Krieger/Bansal (1988) und Dorsch/Teas (1992). Mehta/Moore/Pavia (1992) zeigen darüber hinaus, daß die Operationalisierung inakzeptabler Ausprägungen in der Modellschätzung einen Einfluß auf die Validität der entsprechenden Modelle haben kann.[56] Eine abschließende Beurteilung der verschiedenen Varianten ist auf Basis bisheriger empirischer Befunde folglich nicht möglich.

2.2 Dekompositionelle Verfahren (Conjoint-Analyse)

Grundidee der Conjoint-Analyse ist es, die Probanden zunächst ganzheitlich beschriebene Produktalternativen beurteilen zu lassen. Aufgabe des Marktforschers ist dann, den Nutzenbeitrag, gemessen durch den Teilnutzenwert, jeder einzelnen Eigenschaftsausprägung zum Gesamturteil zu ermitteln.[57] Somit ist die Vorgehensweise der Conjoint-Analyse im wesentlichen dadurch gekennzeichnet, daß statt Bewertungen für die einzelnen Eigenschaften zu einem Gesamturteil zu-

---

52 Vgl. Agarwal/Green (1991), S. 145.
53 Vgl. Green/Krieger/Agarwal (1993), S. 378.
54 Vgl. Srinivasan (1988), S. 304.
55 Zitiert nach Dorsch/Teas (1992), S. 39.
56 Vgl. Mehta/Moore/Pavia (1992), S. 473 f.
57 Vgl. Green/Srinivasan (1978), S. 104.

sammengefügt, Gesamturteile in die einzelnen Nutzenbeiträge der Eigenschaften zerlegt werden.[58] Grundlage hierfür ist wiederum das additive Nutzenmodell:[59]

$$(2\text{-}2) \qquad U_{im} = \sum_{j \in J} \sum_{k \in K_j} \beta_{ijk} \cdot x_{jkm} + \varepsilon_{im} \qquad (i \in I;\ m \in M)$$

wobei:

U_{im}: Geschätzter Gesamtnutzen für die m-te Alternative für den i-ten Probanden,

β_{ijk}: Teilnutzenwert für die k-te Ausprägung der j-ten Eigenschaft für den i-ten Probanden,

$$x_{jkm} = \begin{cases} 1: & \text{wenn bei der m-ten Alternative die k-te Ausprägung der j-ten Eigenschaft auftritt,} \\ 0: & \text{sonst,} \end{cases}$$

ε_{im}: Störterm.

Der entscheidende Unterschied zu den Self-Explicated-Verfahren liegt in der Bestimmung der Parameter dieses Modells.[60] Auf Basis erhobener Gesamturteile für ganzheitlich beschriebene Produktalternativen werden die Teilnutzenwerte β_{ijk} jetzt für die verschiedenen Eigenschaftsausprägungen geschätzt. Aus diesem Grund muß das Modell um einen Störterm ε_{im} erweitert werden.[61] Während diese dekompositionelle Vorgehensweise Grundgedanke aller Conjoint-Verfahren ist, liefert die Literatur zahlreiche Varianten zur Ausgestaltung einer Conjoint-Analyse, die im folgenden dargestellt werden.

2.2.1 Systematisierung dekompositioneller Ansätze zur Präferenzmessung

Die **Conjoint-Analyse** stellt den bedeutendsten Vertreter dekompositioneller Verfahren zur Präferenzmessung dar und steht im Mittelpunkt der folgenden Be-

---

58 Vgl. Stallmeier (1993), S. 11.

59 Es sei darauf hingewiesen, daß auch an dieser Stelle die Annahme eines additiven Modells nicht zwingend ist. Prinzipiell sind ebenso nicht-lineare Modelle einsetzbar. Das additive Modell wird jedoch vor allem aufgrund seiner Robustheit und einfachen Handhabung bei der Schätzung am häufigsten angewendet (vgl. u.a. Backhaus et al. (1996), S. 509 und Reiners (1996), S. 59).

60 Vgl. Jain/Mahajan/Malhotra (1979), S. 249.

61 Zur formalen Darstellung des additiven Modells im Rahmen der Conjoint-Analyse vgl. auch Vriens (1995), S. 9.

trachtung.[62] Die verschiedenen Varianten der Conjoint-Analyse werden kurz vorgestellt, wobei der Schwerpunkt der Ausführungen auf der Vorgehensweise bei der Datenerhebung liegt. Die Varianten werden dabei zunächst nach der **Anzahl simultan zu beurteilender Alternativen** systematisiert. Unterschieden werden dabei folgende 3 Fälle:

1. **alle** zu beurteilenden Stimuli[63] sind simultan zu bewerten,
2. verschiedene **Teilmengen** sind separat zu bewerten oder
3. lediglich **2** Stimuli werden simultan präsentiert.

Teilweise bereits durch die Anzahl der zu beurteilenden Stimuli determiniert, können die verschiedenen Conjoint-Varianten darüber hinaus nach der **Art der Beurteilungsaufgabe** für den Probanden differenziert werden.[64] Diese beiden Abgrenzungskriterien definieren gleichzeitig das Skalenniveau der abhängigen Variable, die in das unterstellte Nutzenmodell eingeht.[65] Abbildung 2-3 gibt einen Überblick über verschiedene Conjoint-Varianten systematisiert nach den beiden angeführten Abgrenzungskriterien.

---

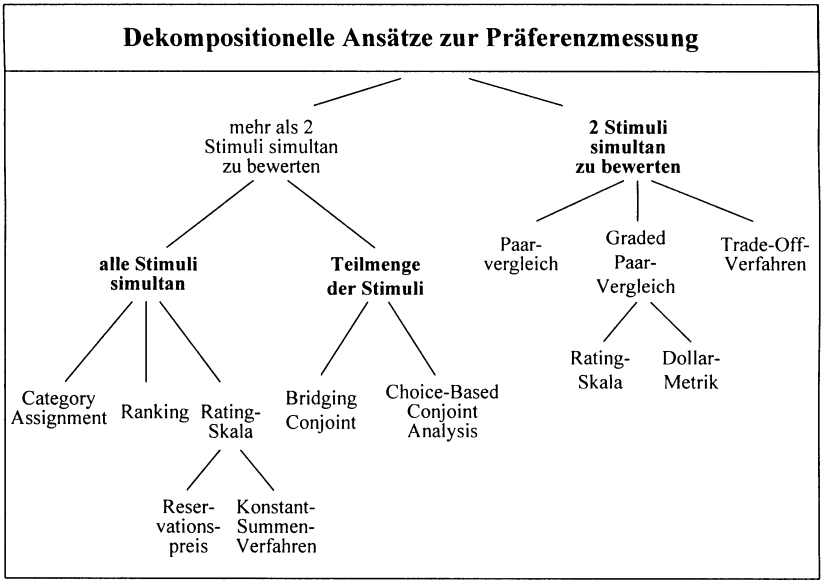
62 Der zweite große Bereich dekompositioneller Verfahren wird unter dem Begriff Multidimensionale Skalierung (MDS) subsumiert (vgl. Herrmann/Schmidt-Gallas/Huber (2000), S. 255). Die Besonderheit der MDS im Vergleich zur Conjoint-Analyse besteht darin, daß die Eigenschaften der zu beurteilenden Produktkonzepte nicht a priori festgelegt werden müssen, sondern nachträglich auf Basis subjektiver Ähnlichkeitsurteile abgeleitet werden (vgl. Backhaus et al. (1996), S. 433.

63 Die im Rahmen einer Conjoint-Analyse zu beurteilenden Produktbeschreibungen werden üblicherweise als Stimuli bezeichnet.

64 Eine alternative Systematisierung der Bewertungsformen in "qualitative" und quantitative" Ansätze schlägt Stallmeier (1993) S. 57 ff. vor.

65 Vgl. hierzu die Ablaufschritte einer Conjoint-Analyse nach Green/Srinivasan (1978) S. 105 bzw. (1990) S. 5.

Abbildung 2-3: Systematisierung dekompositioneller Ansätze



Quelle: Eigene Erstellung

2.2.1.1 Ansätze mit simultaner Bewertung aller Stimuli

In der traditionellen Form der Conjoint-Analyse werden den Probanden **alle** zu beurteilenden Stimuli gleichzeitig vorgelegt. Die Stimuli stehen dabei für Produktalternativen, die bei der sog. Voll-Profil-Methode durch sämtliche zu untersuchenden Produkteigenschaften beschrieben sind.[66] Ohne späteren Ausführungen vorgreifen zu wollen, sei an dieser Stelle bereits angemerkt, daß dieser klassischen Form der Conjoint-Analyse in mehrfacher Hinsicht Grenzen gesetzt sind. Zum einen ist die Anzahl an Eigenschaften, durch die ein Stimulus beschrieben werden kann, speziell bei der Voll-Profil-Methode begrenzt. Green/Srinivasan (1990) empfehlen, nicht mehr als 6 Eigenschaften zu verwenden.[67] Zum anderen kann den Probanden nur eine begrenzte Anzahl an Stimuli insgesamt zur Beur-

---

66 Bei der Präsentationsform der Produktbeschreibungen kann zwischen verbaler, bildlicher und multimedialer Darstellung variiert werden. Des weiteren können den Probanden reale Produkte zur Beurteilung vorgelegt werden. Einen Überblick zu empirischen Untersuchungen, die den Einfluß der Präsentationsform auf die Validität von Conjoint-Analysen überprüfen, geben Ernst/Sattler (2000). Die Variationsmöglichkeiten bei der Präsentationsform werden im Rahmen dieser Arbeit nicht weiter berücksichtigt, da sie für die hier behandelte Fragestellung keine entscheidende Rolle spielen.

67 Vgl. Green/Srinivasan (1990), S. 8.

teilung vorgelegt werden. Hierbei gilt es, in Abhängigkeit der zu schätzenden Parameter ausreichend viele Datenpunkte in Form bewerteter Stimuli zu erheben und gleichzeitig eine Obergrenze von ca. 30 Stimuli nicht zu überschreiten, um die Probanden kognitiv nicht zu überfordern.[68] Diese Obergrenze wird allerdings bereits bei einer sehr geringen Anzahl an Eigenschaften und Ausprägungen erreicht. So besteht ein vollständiges Design bei 3 Eigenschaften mit je 3 Ausprägungen aus 3*3*3=27 Stimuli.[69] Da bei den meisten Anwendungen der Conjoint-Analyse wesentlich mehr Eigenschaften von Interesse für den Anwender sind, werden meist fraktionierte Designs gebildet, die das vollständige Design möglichst gut abbilden.[70] Einen Lösungsvorschlag, um bereits die Anzahl zu betrachtender Eigenschaften und die daraus resultierenden Stimuli zu reduzieren, liefert Schweikl (1985). Im Rahmen eines computergestützten Interviews werden für jeden Probanden die individuell wichtigsten Eigenschaften identifiziert und nur diese im folgenden Conjoint-Interview berücksichtigt.[71] Damit wird die Anzahl zu untersuchender Eigenschaften reduziert und gleichzeitig sichergestellt, daß keine für die Testperson auswahlentscheidenden Attribute vernachlässigt werden.[72] Köcher (1997) schlägt mit der MaiK-Conjoint-Analyse zusätzlich eine Individualisierung der Produktkonzepte vor, zu denen die Eigenschaftsausprägungen kombiniert werden.[73]

Neben der Anzahl simultan zu beurteilender Stimuli können die Varianten der Conjoint-Analyse nach Art der Beurteilungsaufgabe abgegrenzt werden.[74] Die für die Probanden wohl einfachste Form ihre Präferenzen auszudrücken, stellt das sogenannte **Category Assignment** dar. Die Aufgabe für die Probanden besteht

---

68 Green/Srinivasan (1978), S. 109 nennen als Obergrenze 30 Stimuli, um die Probanden kognitiv nicht zu überlasten. Louviere/Gaeth (1988), S. 60 nennen mit 32 Stimuli eine entsprechende Größenordnung.

69 Ein vollständiges Design umfaßt alle konstruierbaren Eigenschaftskombinationen bei gegebener Anzahl an Eigenschaften und deren Ausprägungen.

70 Da die Gestaltung und Fraktionierung der Designs jedoch nicht Kern der vorliegenden Arbeit sind, wird zu dieser Problematik auf die Literatur verwiesen. So führt Stallmeier (1993) beispielsweise explizit eine Simulationsanalyse zum Einfluß der Untersuchungsdesigns auf die Ergebnisstabilität der Conjoint-Analyse durch.

71 Vgl. Schweikl (1985), hier: S. 111.

72 Vgl. Boecker/Schweikl (1988), S. 15.

73 Vgl. Köcher (1997), S. 141.

74 Da sowohl Stallmeier (1993), S. 57-63 als auch Reiners (1996), S. 88-98 einen sehr umfassenden Überblick über Formen der Beurteilungsaufgabe geben, wird an dieser Stelle auf eine ausführliche Diskussion verzichtet und die angegebene Literatur empfohlen.

dabei lediglich in der Zuordnung der Produktkonzepte zu verschiedenen Kategorien. Louviere/Fox/Moore (1993) integrieren in einem Vergleich unterschiedlicher Beurteilungsaufgaben u.a. die Aufgabe, die vorgelegten Stimuli in die Kategorien "würde ich in Betracht ziehen" und "würde ich nicht in Betracht ziehen" einzuordnen.[75] Nicht zuletzt der extrem hohe Informationsverlust durch diese Vorgehensweise ist ein Grund, daß das Category Assignment in der praktischen Anwendung kaum von Bedeutung ist. Vielmehr wird es in Kombination mit anderen Bewertungsformen der Stimuli eingesetzt. So kann einer differenzierten Beurteilung der Stimuli eine Einteilung in verschiedene Kategorien zur Vereinfachung vorangestellt werden.[76] In der klassischen Form der Conjoint-Analyse werden Präferenzurteile durch **Rangreihung (Ranking)** der präsentierten Stimuli erhoben. Der Proband wird gebeten, die ihm vorgelegten Stimuli in die Rangfolge zu sortieren, in der er sie am ehesten wählen bzw. kaufen würde. Sollen die Präferenzunterschiede zwischen den Produktalternativen quantifiziert werden, müssen die Stimuli anhand einer **Rating-Skala** bewertet werden. Dabei kann die Präferenz einer Produktalternative beispielsweise mit einer 11-Punkte-Rating-Skala (1 = "am wenigsten präferiert"; 11 = "am meisten präferiert") konkret[77] oder jeweils in Relation zu einem "Standardprodukt" erfragt werden.[78] Alternativ können die Skalenpunkte einer Rating-Skala auch als Kaufwahrscheinlichkeiten definiert werden. Huber et al. (1993) setzen eine 9-Punkte-Rating-Skala ein, deren Endpunkte als "10% oder niedrigere" bzw. "90% oder höhere" Kaufwahrscheinlichkeit definiert sind.[79] Kalish/Neslin (1991) schlagen weiterhin vor, nach dem **Reservationspreis** für einen Stimuli zu fragen.[80] Somit wird die Präferenz für ein Produkt explizit in Form einer monetären Größe ausgedrückt. Als weitere Bewertungsmöglichkeit sei das **Konstant-Summen-Verfahren** angeführt.[81] Die

---

75 Vgl. Louviere/Fox/Moore (1993), S. 206.

76 Vgl. Reiners (1996), S. 90.

77 Vgl. z.B. Leigh/MacKay/Summers (1984), S. 458. Die Anzahl der Skalenpunkte ist dabei selbstverständlich nicht fest vorgegeben.

78 Vgl. die Definition der Skalenpunkte bei Louviere/Gaeth (1988), S. 68.

79 Vgl. Huber et al. (1993), S. 108.

80 Der Reservationspreis für ein Produkt entspricht dem Preis, den ein Konsument für dieses Produkt maximal zu zahlen bereit ist. Dieser Preis entspricht gleichzeitig dem Nutzen für dieses Produkt bei dem entsprechenden Konsumenten (vgl. Kalish/Nelson (1991) S. 328).

81 Zur Vorgehensweise beim Konstant-Summen-Verfahren vgl. z.B. Mahajan/Green/Goldberg (1982), S. 337.

beiden letztgenannten Methoden können im weiteren Sinne als unterschiedliche Gestaltungsform einer Rating-Skala interpretiert werden.

Welcher Bewertungsform bei der Ermittlung von Nachfragerpräferenzen der Vorzug gegeben werden sollte, ist a priori nicht zu beantworten. Für den Einsatz der Rating-Methode sprechen zweifelsohne die detaillierten Informationen, die durch das implizit höhere Skalenniveau erhoben werden. Jedoch birgt dieses Vorgehen die Gefahr, daß insbesondere bei einer großen Anzahl an Stimuli diese den Probanden zwar simultan vorgelegt, tatsächlich aber separat beurteilt werden. Demgegenüber werden die Testpersonen bei einer Ranking-Aufgabe zu tatsächlichem Abwägen zwischen den verschiedenen Eigenschaften gezwungen, was der eigentlichen Idee der Conjoint-Analyse wesentlich näher kommt.[82] Als Kritikpunkt sowohl an der Bewertung der Stimuli durch Rangreihung als auch durch eine (beliebige) Rating-Skala wird immer wieder angeführt, daß lediglich Präferenzen, nicht aber tatsächliches Auswahlverhalten von Konsumenten erhoben werden.[83] Malhotra (1986) schlägt deshalb vor, die Produktalternativen zunächst danach zu differenzieren, ob grundsätzlich eine Kaufbereitschaft besteht. Sofern Produkte völlig inakzeptabel für eine Testperson sind, werden für diese keine Aussagen zu deren Präferenz mehr erfragt. Präferenzen werden in einem zweiten Schritt lediglich für die Alternativen erhoben, die für den Befragten in Frage kommen.[84] Auf dem gleichen Grundgedanken basiert die von Voeth/Hahn (1998) vorgestellte Limit Conjoint-Analyse. Sie stellt ebenso eine Erweiterung der traditionellen Conjoint-Analyse dar, bei der zusätzlich zu den erhobenen Präferenzurteilen eine Auswahlentscheidung integriert wird. Damit sollen zusätzlich zu den Präferenzen gegenüber den Stimuli diejenigen identifiziert werden, für die die Probanden auch eine Kaufbereitschaft aufweisen.[85] Im Gegensatz zu Malhotra (1986) wird die Kaufbereitschaft im Anschluß an die erhobenen Präferenzen abgefragt.

---

82 Vgl. Stallmeier (1993), S. 57.

83 Vgl. hierzu Louviere/Gaeth (1988), S. 59; Oliphant et. al. (1992), S. 385 und Elrod/Louviere/Davey (1992), S. 368.

84 Vgl. Malhotra (1986), S. 34. Die Vorgehensweise von Malhotra stellt ein Beispiel für das oben beschriebene Category Assignment dar. Eine Besonderheit liegt darin, daß für die Stimuli, die für einen Kauf überhaupt nicht in Frage kommen, keine weiteren Präferenzurteile erhoben werden.

85 Vgl. Voeth/Hahn (1998), S. 120 f.

2.2.1.2 Ansätze mit simultaner Bewertung einer Teilmenge der Stimuli

Um eine Informationsüberlastung der Probanden zu vermeiden und gleichzeitig eine größere Anzahl an Eigenschaften berücksichtigen zu können, besteht die Möglichkeit, die zu beurteilenden Stimuli auf einzelne **Teilmengen** zu verteilen. Bei sogenannten "Bridging"-Designs werden alle Stimuli auf verschiedene Stapel verteilt, die jeweils separat beurteilt werden. Sichergestellt werden muß lediglich, daß wenigstens eine Eigenschaft bei allen Produktbeschreibungen identisch ist, um die Ergebnisse der einzelnen Teilmengen für die Schätzung wieder zusammenfügen zu können.[86] Die Reduzierung der gleichzeitig zu beurteilenden Stimuli mittels "Bridging" ist unabhängig von der Art der Beurteilungsaufgabe und kann prinzipiell mit allen diskutierten Formen der Präferenzabfrage kombiniert werden. Obwohl Bretton-Clark`s BRIDGER eine Software zur Gestaltung eines Bridging-Designs anbietet, ist diese Vorgehensweise bisher verhältnismäßig selten eingesetzt worden.[87]

Eine wesentlich bedeutendere Rolle spielt die Aufteilung der zu beurteilenden Stimuli auf verschiedene Teilmengen in Zusammenhang mit der sog. **Choice-Based Conjoint Analysis (CBC).**[88] Neben der Reduzierung simultan zu bewertender Stimuli ist die CBC maßgeblich dadurch gekennzeichnet, daß den Probanden tatsächliche Auswahlentscheidungen abverlangt werden. Hierfür wird jeweils eine **Teilmenge** an Stimuli (sog. Choice-Set) präsentiert, aus denen sie lediglich die für sie beste Alternative auswählen. Auf diese Weise werden die Probanden zu einer echten Auswahlentscheidung gezwungen, statt Präferenzen zwischen Alternativen anzugeben. Damit sind bei der CBC im Gegensatz zu Ranking- oder Rating-basierten Conjoint-Analysen keine Annahmen über das Kaufverhalten zu treffen, da das Auswahlverhalten direkt beobachtet wird.[89] Eine weitere Besonderheit der CBC liegt in der Berücksichtigung einer "Nicht-Kauf-Alternative" in jedem Choice-Set. Damit haben die Probanden die Möglichkeit, alle innerhalb eines Choice-Sets präsentierten Stimuli abzulehnen, sofern keine für sie attraktive

---

86 Vgl. Green/Srinivasan (1990), S. 9.

87 Nach Kenntnis der Autorin sind Studien von Rosko et al. (1985) und Baalbaki/Malhotra (1995) die einzigen Veröffentlichungen, in denen ein Bridging-Design angewendet wurde.

88 Zu den Grundlagen der Choice-Based Conjoint Analysis vgl. Batsell/Louviere (1991); Louviere (1988) und Louviere/Woodworth (1983).

89 Vgl. Weiber/Rosendahl (1997), S. 109.

Alternative angeboten wird.[90] Insgesamt wird der CBC ein höherer Realitätsbezug zugesprochen, da die Beurteilungsaufgabe wesentlich näher an der tatsächlichen Kaufsituation liegt und damit besser geeignet ist, das Kauf- bzw. Auswahlverhalten zu prognostizieren.[91]

Ein wesentlicher Nachteil der CBC betrifft allerdings das Aggregationsniveau der geschätzten Teilnutzenwerte. Während die traditionelle Conjoint-Analyse die Schätzung individueller Nutzenfunktionen ermöglicht, lassen Choice-Experimente lediglich eine aggregierte (ggf. segmentweise) Auswertung zu.[92] Auch wenn beispielsweise Oliphant et. al. (1992) zeigen können, daß die Ergebnisse einer CBC nachträglich disaggregiert werden können, um individuelle Unterschiede in den Teilnutzenwerten berücksichtigen zu können[93], zielen neuere Auswertungsmethoden darauf ab, die Heterogenität zwischen den Probanden bereits bei der Schätzung einzubeziehen.[94] Dieses Vorgehen stellt einen weiteren Versuch dar, individuelle Unterschiede in die auf der Basis von Choice-Experimenten ermittelten Nutzenfunktionen zu integrieren. Ein weiterer Kritikpunkt an der CBC betrifft die Ineffizienz, mit der Präferenzdaten erhoben werden. Während vom Probanden jeweils nur die für ihn beste Alternative eines Choice-Sets identifiziert wird, muß der Proband für diese Auswahlentscheidung eine Vielzahl an Informationen erfassen und verarbeiten. Jeder Stimulus wird durch mehrere Eigenschaften beschrieben und jedes Choice-Set besteht aus mehreren Produktalternativen.[95] Durch die sehr eingeschränkte Abfrage nach der individuell besten Alternative je Choice-Set, werden nur sehr wenige Informationen letztendlich erhoben. Der Informationsverlust wird noch größer, wenn der Proband die "Nicht-Kauf-Alternative" wählt, da in diesem Fall keinerlei Aussagen zu seinen Präferenzen zwischen den einzelnen Alternativen möglich sind.[96]

---

90 Vgl. Oliphant et. al. (1992), S. 385.

91 Zu den Vorteilen von Choice-Experimenten vgl. z.B. Louviere/Gaeth (1988), S. 62.

92 Vgl. u.a. Louviere (1988), S. 99 und Elrod/Louviere/Davey (1992), S. 368.

93 Vgl. Oliphant et al. (1992), S. 393.

94 Neuere Ansätze zur Berücksichtigung von Heterogenität in Choice-Experimenten werden u.a. von DeSarbo/Ramaswamy/Cohen (1995); Johnson (1997b) und Zwerina/Huber (1996) vorgeschlagen.

95 Vgl. Sawtooth Software (1995), S. 1.

96 Vgl. Elrod/Louviere/Davey (1992), S. 370.

2.2.1.3 Ansätze mit simultaner Bewertung von genau zwei Stimuli

Die kleinste Teilmenge und damit die Untergrenze simultan zu bewertender Stimuli liegt bei zwei. In der einfachen Form der daraus resultierenden **Paarvergleiche** gibt die Testperson lediglich an, welchen der beiden Stimuli sie vorzieht. Darüber hinaus muß bei einem sogenannten **"Graded Paired Comparison"** zusätzlich angegeben werden, wie stark der entsprechende Stimulus bevorzugt wird. Das Ausmaß der Präferenz kann durch den Einsatz einer Rating-Skala quantifiziert werden oder die Testperson gibt im Rahmen einer "Dollar-Metrik" an, wieviel mehr sie bereit wäre, für die bevorzugte Produktalternative zu bezahlen.[97] Für den Einsatz von Paarvergleichen spricht die Tatsache, daß die Anzahl zu verarbeitender Informationen pro Beurteilungsaufgabe minimiert wird und Präferenzurteile zwischen den Paaren in eine Rangfolge aller Stimuli transformiert werden können.[98] Gleichzeitig darf dieser Vorteil allerdings nicht darüber hinwegtäuschen, daß gerade diese geringe Zahl an Informationen je Paarvergleich eine häufige Wiederholung derselben mit unterschiedlichen Stimuli erforderlich macht. Veranschaulicht werden kann dies daran, daß für die Ableitung einer Rangfolge für n Stimuli insgesamt $n*(n-1)/2$ Paarvergleiche erforderlich sind. Dies macht die Ineffizienz der Datenerhebung auf Basis von Paarvergleichen deutlich.[99]

Als alternatives Vorgehen zur Voll-Profil-Methode wird häufig in der Literatur die von Johnson (1974) vorgestellte **Trade-Off-Methode** genannt.[100] Die Besonderheit dieser Methode liegt darin, daß die zu bewertenden Stimuli nicht durch sämtliche Eigenschaften beschrieben werden, sondern lediglich durch zwei Eigenschaften.[101] Die Aufgabe der Probanden besteht darin, die Kombination verschiedener Ausprägungen zweier Eigenschaften, dargestellt in einer Trade-Off-Matrix, zu bewerten. Damit kann diese Vorgehensweise, die auch als **Two-**

---

97 Vgl. z.B. Leigh/MacKay/Summers (1984), S. 457 oder Stallmeier (1993), S. 221.

98 Vgl. Reynolds/Jolly (1980), S. 532.

99 Vgl. hierzu auch Reiners (1996), S. 93 und Louviere/Gaeth (1988), S. 63.

100 Zur Trade-Off-Methode vgl. insbesondere Johnson (1974).

101 Es sei darauf hingewiesen, daß auch bei den bereits diskutierten Varianten der Conjoint-Analyse nicht zwingend die Voll-Profil-Methode, d.h. die Beschreibung der Stimuli anhand sämtlicher Eigenschaften, eingesetzt werden muß. Auch hier ist eine Reduzierung der Eigenschaften prinzipiell möglich. Die Trade-Off-Methode ist dagegen dadurch gekennzeichnet, daß die Stimuli immer durch *genau* zwei Eigenschaften beschrieben werden.

factor-at-a-time-Verfahren bezeichnet wird, nur eingeschränkt der gewählten Systematisierung zugeordnet werden. Statt zwei simultan zu beurteilender Stimuli, sind es *genau* zwei Eigenschaften, durch die Alternativen beschrieben werden und zwischen denen ein Trade-Off gemacht werden muß. Diese Besonderheit ist jedoch maßgebliches Kennzeichen der Methode und ist ausschlaggebend für deren Einordnung an dieser Stelle. Analog zu Paarvergleichen mit Voll-Profilen liegt der Vorteil dieser Methode in dem begrenzten Maß simultan zu verarbeitender Informationen. Damit verbunden ist aber ebenfalls eine gewisse Ineffizienz bei der Datenerhebung. Zudem kann als Kritikpunkt angeführt werden, daß die Trade-Off-Analyse der dekompositionellen Grundidee der Conjoint-Analyse nur bedingt gerecht wird, da statt ganzheitlicher Produkte vereinfachte Eigenschaftskombinationen bewertet werden. Der damit verbundene Realitätsverlust der Befragungsform als auch die Ineffizienz in der Datenerhebung mögen Gründe dafür sein, daß die Trade-Off-Methode in der praktischen Anwendung eine deutlich geringere Bedeutung als der Voll-Profil-Methode zukommt.[102]

2.2.2 Empirische Vergleiche zwischen unterschiedlichen dekompositionellen Ansätzen

Die Ungewißheit darüber, mit welcher Form der Conjoint-Analyse die tatsächlichen Präferenzen von Konsumenten am ehesten erfaßt werden können, haben zu einer intensiven Diskussion der verschiedenen Varianten in der Literatur geführt. So wundert es nicht, daß zahlreiche empirische Vergleiche durchgeführt wurden. Die bisherigen Forschungsergebnisse werden an dieser Stelle zusammengefaßt, um zu prüfen, ob sich in diesen Untersuchungen einzelne Formen der Conjoint-Analyse als überlegen ausgezeichnet haben. Die verschiedenen Studien unterscheiden sich dabei sowohl in den jeweils betrachteten Methoden als auch hinsichtlich der eingesetzten Gütemaße zur Beurteilung der Methoden. Da die Definition und Abgrenzung der verschiedenen Beurteilungskriterien in der Literatur

---

[102] Wittink/Cattin (1989) berichten, daß bei kommerziellen Conjoint-Studien in den USA in der Zeit von 1971-1980 in 27% aller Studien die Trade-Off-Methode eingesetzt wurde. Von 1981-1985 war dies nur noch in lediglich 6% der Anwendungen der Fall (vgl. Wittink/Cattin (1989), S. 93). Überraschenderweise können Wittink/Vriens/Burhenne (1994) bei einem entsprechenden Überblick mit 15% zwar über einen stärkeren Einsatz der Trade-Off-Methode bei kommerziellen Anwendungen in Europa berichten, jedoch liegt der Anteil auch hier niedriger als bei der Voll-Profil-Methode mit 24% (vgl. Wittink/Vriens/Burhenne (1994), S. 45). Johnson, auf den die Methode zurückgeführt werden kann, spricht selbst davon, daß die Trade-Off-Analyse in der praktischen Anwendung bereits "obsolet" geworden ist (zitiert nach Wittink/Vriens/Burhenne (1994), S. 45).

teilweise voneinander abweichen, wird deren Operationalisierung in dieser Arbeit im folgenden Abschnitt kurz erläutert.[103]

2.2.2.1 Gütemaße

Reliabilität:

Ausgewiesen werden die Studien, die explizit die Reliabilität verschiedener Methoden vergleichen.[104] Grundlage für dieses Gütemaß ist eine wiederholte Befragung derselben Person mit der gleichen Datenerhebungsmethode und den identischen Stimuli. Die Reliabilität mißt dann den Grad der Übereinstimmung der Ergebnisse und läßt eine Aussage bzgl. der Konsistenz im Antwortverhalten der Probanden zu.[105]

Diskriminierende Validität:

Unter diesem Validitätskriterium werden die Vergleichsstudien betrachtet, bei denen die geschätzten Teilnutzenwerte oder daraus ermittelte Bedeutungsgewichte verschiedener Methoden gegenübergestellt werden. Die Frage dabei ist, ob unterschiedliche Methoden zur Erfassung von Präferenzen zu Unterschieden in den geschätzten Präferenzstrukturen führen.[106]

---

103 Zu unterschiedlichen Begriffsabgrenzungen der Gütemaße vgl. Bateson/Reibstein/Boulding (1987).

104 Zum Begriff der Reliabilität vgl. Hammann/Erichson (1994), S. 75 ff. und Böhler (1992), S. 102 ff.

105 Verschiedene Autoren sprechen ebenfalls von der Überprüfung der Reliabilität, sofern die Aufgabe mit derselben Datenerhebungsmethode aber unterschiedlichen Untersuchungsobjekten (Stimuli) wiederholt durchgeführt wird (vgl. u.a. Agarwal/Green (1991), S. 144). Entsprechend dieser weiten Begriffsabgrenzung müßten alle Tests auf Basis eines "Holdout-Sample" unter Reliabilitätsprüfungen subsumiert werden (vgl. Bateson/Reibstein/Boulding (1987), S. 473). Dabei wird allerdings nicht die Korrespondenz eingesetzter Inputdaten überprüft, sondern es werden bereits Prognosen basierend auf den geschätzten Nutzenmodellen durchgeführt. Aus diesem Grund wird eine solche Vorgehensweise (insbesondere mit einem Holdout-Sample) im Rahmen dieser Arbeit als Überprüfung der Prognosegüte definiert.

106 Dabei werden auch die Studien herangezogen, bei denen ursprünglich die konvergierende Validität getestet wird. Eine Überprüfung der konvergierenden Validität impliziert einen entsprechenden Vergleich zwischen den Verfahren. Der Unterschied besteht lediglich in der Formulierung der Ausgangshypothese. In beiden Fällen wird getestet, ob Unterschiede zwischen den Ergebnissen alternativer Verfahren auftreten. Während bei der konvergierenden Validität keine Unterschiede erwartet werden, wird die Ausgangshypothese bei der diskriminierenden Validität genau umgekehrt formuliert. In diesem Fall wird vermutet, daß verschiedene Methoden der Präferenzmessung zu unterschiedlichen Ergebnissen führen. Entsprechend wird an dieser Stelle von diskriminierender Validität gesprochen.

Interne Validität

Die interne Validität stellt ein Gütemaß dar, bei dem die Modellgüte an sich überprüft wird. Gemessen wird die Übereinstimmung zwischen den Inputdaten und den auf deren Grundlage geschätzten Werten für die abhängige Variable.[107] Als abhängige Variable gehen die empirisch beobachteten Präferenzdaten in das Modell ein, die durch die geschätzten Outputdaten rechentechnisch rekonstruiert werden sollen.[108]

Prognosevalidität:

Vornehmliches Ziel der Präferenzmessung ist, auf Basis erhobener Präferenzdaten und den daraus ermittelten Präferenzstrukturen das Wahlverhalten bzw. Kaufverhalten der Probanden auch für alternative und neue Produkte prognostizieren zu können. Mit der Prognosevalidität wird explizit die Eignung der unterschiedlichen Ansätze hierfür getestet. In zahlreichen Studien wird zu diesem Zweck mit Hilfe eines Holdout-Sample[109] eine Kaufsituation simuliert. Die Prognosegüte prüft dann, wie gut mittels der geschätzten Präferenzstrukturen das Verhalten der Probanden in dieser simulierten Kaufentscheidung prognostiziert werden kann.[110]

Externe Validität:

Die externe Validität stellt das stärkste Validitätsmaß dar. Sie liegt dann vor, wenn unter Verwendung der Schätzergebnisse das Entscheidungsverhalten bzw. Präferenzen in von der Befragung völlig unabhängigen Situationen prognostiziert werden kann. Beispiele hierfür sind die Prognose realer Marktanteile oder des tatsächlichen Entscheidungsverhaltens zum Untersuchungsgegenstand, ggf. mit einer zeitlichen Verzögerung.[111] Ein Vergleich zwischen den auf Basis einer Befragung geschätzten und den tatsächlich beobachteten Präferenzen stellt ein Maß für die externe Validität der Schätzergebnisse dar.

---

107 Vgl. Green/Srinivasan (1978), S. 115.

108 Vgl. hierzu auch Baier/Säuberlich (1997), S. 968.

109 Von einem Holdout-Sample wird gesprochen, wenn unabhängig von der eigentlichen Befragungsmethode, Präferenzen für zusätzliche Stimuli erhoben werden. Die so erhobenen Präferenzdaten gehen nicht in die Schätzung der Nutzenwerte ein, sondern werden ausschließlich zur Validitätsprüfung herangezogen.

110 Green/Srinivasan (1978) sprechen in diesem Zusammenhang von Kreuzvalidierung (vgl. S. 115).

111 Vgl. hierzu auch Green/Srinivasan (1990), S. 13.

2.2.2.2 Bisherige empirische Vergleiche

Tabelle 2-1 faßt die bisherigen Ergebnisse empirischer Vergleiche verschiedener Varianten der Conjoint-Analyse zusammen. Die Studien sind dabei nach den jeweils untersuchten Methoden und den eingesetzten Gütemaßen gegliedert. Die Systematisierung der Methoden nach Anzahl simultan zu beurteilender Stimuli wird auch an dieser Stelle beibehalten und die Tabelle in Anlehnung daran untergliedert. Die Hauptergebnisse der Studien werden jeweils zusammengefaßt, wobei Aussagen zur Signifikanz auf einem Signifikanzniveau von mindestens 10% (d.h. $p<0,1$) basieren. Teilweise werden die Gütemaße in den Untersuchungen auf verschiedene Arten operationalisiert. Sofern sich die Ergebnisse alternativer Operationalisierungen hinsichtlich eines Kriteriums widersprechen, kann kein eindeutiges Ergebnis abgeleitet werden. Dies wird bei der Ergebnisinterpretation berücksichtigt und innerhalb der Tabelle dokumentiert. Des weiteren sind in Tabelle 2-1 der Untersuchungsgegenstand, die Anzahl der betrachteten Eigenschaften und der Stichprobenumfang der Studie angegeben, um mögliche systematische Einflüsse auf die Ergebnisse identifizieren zu können. Die letzte Spalte der Tabelle gibt Auskunft darüber, ob für die verschiedenen zu vergleichenden Methoden separate Stichproben befragt wurden (between-subject-design) oder ob die Probanden mehrere in der Studie untersuchten Methoden beantwortet haben (within-subject-design). Gegen ein solches within-subject-design spricht, daß die Ergebnisse sowohl durch Lern- als auch Reihenfolgeeffekte verzerrt werden können.[112]

---

[112] Vgl. z.B. Agarwal/Green (1991), S. 145, Huber et al. (1993), S. 110 und Johnson (1991), S. 223. Angemerkt sei, daß die Problematik des Reihenfolgeeffekts bei einen within-subject Design an Gewicht verliert, sofern zwischen den einzelnen Befragungsschritten eine ausreichend lange Zeitverzögerung auftritt (vgl. z.B. Huber/Herrmann/Gustafsson (2000), S. 191).

Tabelle 2-1: Empirische Vergleichsstudien dekompositioneller Verfahren zur Präferenzmessung (Teil 1)

Studie	Verglichene Methoden	Gütemaße	Ergebnisse	Produkte	Anzahl Eigenschaften	Stichprobe	Experimentdesign
"Alle Stimuli"							
Huber/Herrmann/ Gustafsson (2000)	Ranking Rating Ranking+Rating	Diskriminierende Validität	keine sig. Unterschiede c)	Fernseher	5	154	within
Kalish/Nelson (1991)	Ranking Rating Reservation Price	Prognosevalidität	Ranking und Rating sig. besser als Reservation Price kein sig. Unterschied zwischen Ranking und Rating	Flugticket	3	255 b)	between
"Alle Stimuli" vs. "Teilmenge von Stimuli"							
Louviere/Gaeth (1988)	Rating Choice	Diskriminierende Validität	kein sig. Unterschied	Flugtickets Berufswahl Stadtpark	4 22 5	35 400 b) 512	within between within
Louviere/Fox/ Moore (1993)	Rating Choice	Diskriminierende Validität	kein sig. Unterschied zwischen Choice-Varianten sig. Unterschied zwischen Rating und Choice	Reiseziele	9	168	within
Oliphant et al. (1992)	Rating Choice	Diskriminierende Validität Prognosevalidität	sig. Unterschied kein Unterschied a)	Versicherungsangebot	10	300	between
Elrod/Louviere/ Davey (1992)	Rating Choice	Prognosevalidität	kein Unterschied a)	Appartment	4	115 b)	within
Moore/Gray-Lee/ Louviere (1998)	Rating Choice	Prognosevalidität	variiert über Kriterien	Zahncreme	5	184 b)	within

Tabelle 2-1: Empirische Vergleichsstudien dekompositioneller Verfahren zur Präferenzmessung (Teil 2)

Studie	**Verglichene Methoden**	**Gütemaße**	**Ergebnisse**	**Produkte**	**Anzahl Eigen-schaften**	**Stich-probe**	**Experi-ment-design**
Scott/Keiser (1984)	Ranking Rating Choice	Diskriminierende Validität Externe Validität	kein sig. Unterschied zwischen Ranking und Choice Choice besser als Rating a)	Solar-Ener-gie-Systeme	10 18	5 4	within within
"Alle Stimuli" vs. "2 Stimuli"							
Reynolds/Jolly (1980)	Ranking Rating Paired Comparison	Reliabilität	Ranking sig. besser Rating Paired Comparison sig. besser als Rating kein sig. Unterschied zwischen Ranking und Paired Comparison	Wertvor-stellungen	12 und 18	54 b)	between
Leigh/MacKay/ Summers (1984)	Ranking Rating Paired comparison Graded Paired Comparison Dollar Metrik	Reliabilität Prognosevalidität	Keine sig. Unterschiede hinsichtlich beider Kriterien zwischen allen Varianten	Taschen-rechner	5	122 b)	between
Reibstein/Bateson/ Boulding (1988)	Rating Paired Comparison Trade-Off	Reliabilität	Kein sig. Unterschied zwischen Paired Comparison und Trade-Off Trade-Off sig. besser als Rating	Telefon-service, Schreib-maschine, Joghurt, Bankservice, Fernseher	5 oder 6	600	within
Segal (1982)	Ranking Trade-Off	Reliabilität	Ranking sig. besser als Trade-Off	Appartment	3	50 b)	between

Tabelle 2-1: Empirische Vergleichsstudien dekompositioneller Verfahren zur Präferenzmessung (Teil 3)

Studie	Verglichene Methoden	Gütemaße	Ergebnisse	Produkte	Anzahl Eigenschaften	Stichprobe	Experimentdesign
Oppedijk van Veen/ Beazley (1977)	Ranking Trade-Off	Diskriminierende Validität	kein Unterschied a)	Langlebiges Konsumgut	3	600	between
Müller-Hagedorn/ Sewing/Toporowski (1993)	Ranking Trade-Off	Diskriminierende Validität	sig. Unterschied	Fernseher Erfrischungsgetränk	5 5	120 50	within within
Jain/Acito/ Malhotra/Mahajan (1979)	Ranking Trade-Off	Prognosevalidität	kein sig. Unterschied	Bankkonto	5	212	between
Rosko/McKenna (1983)	Ranking Trade-Off	Prognosevalidität	Ranking besser als Trade-Off a)	Ambulanter Gesundheitsservice	8	97 b)	between
Safizadeh (1989)	Ranking Trade-Off	Prognosevalidität	Ranking sig. besser als Trade-Off	Krankenhaus	6	107 b)	within
"Alle Stimuli" vs. "Teilmenge von Stimuli" vs. "2 Stimuli"							
Stallmeier (1993)	Rating Choice Paired Comparison Graded Paired Comparison Dollar Metrik	Diskriminierende Validität	Unterschiede bestehen zwischen Rating, Graded Paired Comparison und Choice a) Keine Unterschiede treten zwischen Paired Comparison; Graded Paired Comparison und Dollar Metrik auf	Kurortmarketing	3 und 5	268	between und within

a) keine Aussage zur Signifikanz; b) Befragt wurden Studierende
c) Die Autoren berichten vereinzelt signifikante Unterschiede zwischen Eigenschaftsausprägungen
Quelle: Eigene Erstellung in teilweiser Anlehnung an Sattler/Hensel-Börner (2000), S. 128 f.

Der vorliegende Literaturüberblick zeigt insbesondere im Hinblick auf die angeführten Argumente für die jeweiligen Methoden folgende zentrale Ergebnisse:

1. Der Einsatz von Rating-Skalen bei der Abfrage von Präferenzen dominiert insbesondere in den neueren Studien relativ stark gegenüber der Anwendung einfacherer Ranking-Skalen. Die Ranking-Methode taucht hauptsächlich in Vergleichen mit dem Trade-Off-Verfahren auf. Ein Hauptargument, der Rating-Methode den Vorzug zu geben, liegt in dem höheren Informationsgehalt und dem damit verbundenen Skalenniveau der Rating-Daten. Die darin vermutete Überlegenheit der Methode kann jedoch mit den vorliegenden empirischen Ergebnissen nicht eindeutig bestätigt werden. Während sowohl Kalish/Nelson (1991) als auch Leigh/MacKay/Summers (1984) keine signifikanten Unterschiede in der Prognosevalidität zwischen der Ranking- und Rating-Variante nachweisen können, zeigen die Ergebnisse von Reynolds/Jolly (1980) auf eine signifikant höhere Reliabilität für die Ranking-Methode. Auch innerhalb der Gruppe mit nur 2 simultan zu beurteilenden Stimuli kann Stallmeier (1993) zeigen, daß die Vereinfachung von Paarvergleichen zu keinem nennenswerten Informationsverlust führt. Zum gleichen Ergebnis kommen Leigh/MacKay/Summers (1984).
2. Die Choice-Based Conjoint Analysis wurde bisher fast ausschließlich mit der Rating-Methode aus der Gruppe "alle Stimuli" verglichen. Zwar treten (teilweise signifikante) Unterschiede im Hinblick auf die diskriminierende Validität zwischen den beiden Varianten Rating und Choice auf, aber alle Ergebnisse zur Prognosevalidität deuten darauf hin, daß sich diese Unterschiede in den geschätzten Präferenzstrukturen nicht auf die Prognosegüte der Verfahren auswirken. Lediglich Scott/Keiser (1984) berichten, daß die Choice-Variante hinsichtlich der externen Validität den Rating-Daten überlegen ist. Ob diese auch statistisch signifikant ist, wird nicht dokumentiert, zumal die Autoren nur eine sehr kleine Stichprobe befragt haben. Das Argument für die Choice-Variante, durch direkt beobachtetes Auswahlverhalten solches auch besser prognostizieren zu können, wird durch neuere empirische Ergebnisse nicht gefestigt.
3. Ob die Trade-Off-Methode den dekompositionellen Verfahren im eigentlichen Sinn unterlegen ist, kann vor dem Hintergrund der bisherigen Ergebnisse abschließend nicht beurteilt werden. Zwar liegen eine Reihe von Vergleichsstudien (vor allem mit der Ranking-Methode) vor, jedoch sind die Ergebnisse

nicht eindeutig. So schneidet die (Voll-Profil-) Ranking-Methode in 3 Studien besser ab, gleichzeitig treten aber in entsprechend vielen Fällen keine Unterschiede auf. Die Rating-Variante unterliegt sogar der Trade-Off-Methode bei dem einzigen Vergleich der beiden Methoden. Der stark zurückgehende Einsatz der Trade-Off-Methode kann daher nicht eindeutig auf eine empirisch begründete Unterlegenheit zurückgeführt werden. Vielmehr kann davon ausgegangen werden, daß diese traditionelle Form mittlerweile durch die Adaptive Conjoint Analysis (ACA) ersetzt wurde, die auf dem gleichen Prinzip basiert.[113]

Bei der Interpretation der Ergebnisse muß allerdings eingeräumt werden, daß diese Hauptergebnisse nur als Tendenzen anzusehen sind. Ein abschließendes Urteil für eine Variante der dekompositionellen Ansätze ist aufgrund dieses Literaturüberblicks nicht zulässig. Vor allem muß berücksichtigt werden, daß die Ergebnisse über die verschiedenen Studien hinweg nur bedingt vergleichbar sind. So sind die verschiedenen Beurteilungskriterien je nach Untersuchungsziel unterschiedlich zu gewichten. Des weiteren differieren die Studien sowohl in der Anzahl zu beurteilender Eigenschaften, dem Untersuchungsgegenstand und der Stichprobe. Der Einfluß dieser Untersuchungsmerkmale auf die Ergebnisse eines Validitätsvergleichs ist nach wie vor unklar und darf nicht unterschätzt werden. So argumentieren beispielsweise Green/Krieger/Agarwal (1993), daß Variablen wie Stimulikonstruktion, Experimentdesign, Anleitungen für die Probanden oder das gewählte Validitätsmaß stärkeren Einfluß auf die Validitätsunterschiede zwischen Methoden ausüben als die untersuchten Methoden selbst.[114]

2.3 Vergleich zwischen der Self-Explicated-Methode und der Conjoint-Analyse

Bisher wurden sowohl die Self-Explicated-Methode als auch die verschiedenen Varianten der Conjoint-Analyse separat betrachtet. Ziel der nachstehenden Ausführungen ist es, diese beiden unterschiedlichen Ansätze zur Präferenzmessung einander gegenüberzustellen. Dabei werden die Self-Explicated-Methode und die Conjoint-Analyse ganzheitlich betrachtet, ohne die verschiedenen Varianten ex-

---

[113] Die Adaptive Conjoint Analysis (ACA) stellt eines der hybriden Verfahren dar, die in Kapitel 2.4 betrachtet werden.

[114] Vgl. Green/Krieger/Agarwal (1993), S. 378.

plizit zu berücksichtigen. Für dieses Vorgehen sprechen zweierlei Gründe. Zum einen werden an dieser Stelle die kompositionellen und dekompositionellen Verfahren grundsätzlich diskutiert und die angeführten Argumente treffen in ihrer Tendenz für alle Varianten des jeweiligen Verfahrens gleichermaßen zu.[115] Der Aussagegehalt des Vergleichs wird folglich nicht maßgeblich eingeschränkt. Zum anderen lassen bisherige empirische Ergebnisse keine eindeutigen Aussagen zu. Weder bei den unterschiedlichen Formen der Self-Explicated-Methode noch bei der Conjoint-Analyse hat sich eine Gestaltungsform als eindeutig überlegen herausgestellt. Das Fehlen eindeutiger signifikanter Unterschiede bei bisherigen Untersuchungen zwischen den Varianten rechtfertigt somit eine zusammengefaßte Betrachtung der beiden Verfahrensbereiche. Die vergleichende Diskussion der beiden Verfahren ist zweigeteilt. Zunächst erfolgt eine Gegenüberstellung anhand der in der Literatur angeführten theoretischen Argumente (vgl. Tabelle 2-2). Eine Zusammenfassung der Ergebnisse bei empirischen Vergleichen der beiden Methoden zeigt dann, ob die theoretischen Vor- und Nachteile auch in der Anwendung in erwarteter Form zum Tragen kommen (vgl. Tabelle 2-3).

2.3.1 Theoretischer Vergleich

Als Grundlage der theoretischen Gegenüberstellung der Conjoint-Analyse und der Self-Explicated-Methode dienen die in der Literatur häufig angeführten Vor- und Nachteile der beiden Ansätze.[116] Bei der folgenden Darstellung werden die Argumente für und gegen ein Verfahren nach verschiedenen Kriterien systematisiert. In einer tabellarischen Aufstellung werden sowohl die Ausprägungen der entsprechenden Kriterien bei jeder Methode aufgeführt, als auch die damit verbundenen positiven (+) oder negativen (–) Auswirkungen (vgl. Tabelle 2-2).

---

[115] Eine Ausnahme bildet hierbei das Trade-Off-Verfahren, das der Grundidee der Conjoint-Analyse nur bedingt gerecht wird. Da die dekompositionelle Vorgehensweise im eigentlichen Sinn Gegenstand des folgenden Vergleichs ist, wird das Trade-Off-Verfahren als Variante der Conjoint-Analyse im Rahmen der theoretischen Diskussion ausgeklammert.

[116] Zu den Vor- und Nachteilen der beiden Verfahren vgl. Green/Srinivasan (1990), S. 9 f.; Vriens (1995), S.7 und Sattler/Hensel-Börner (2000), S. 123 ff. Die folgenden Ausführungen beziehen sich, sofern nicht anders gekennzeichnet, auf diese Literaturangaben.

Tabelle 2-2: Theoretischer Vergleich zwischen der Conjoint-Analyse und der Self-Explicated-Methode

Kriterium	Conjoint-Analyse (dekompositionell)	Self-Explicated-Methode (kompositionell)
Befragungsform	**indirekt** + Bewertung ganzheitlicher Produkte entspricht eher der realen Kaufsituation	**direkt** – Probleme bei der Definition des Konzepts „Wichtigkeit" – Verzerrungen durch sozial erwünschte Antworten
Komplexität	**hoch** – steigt mit zunehmender Anzahl an Eigenschaften – erfordert größeres Methodenwissen vom Marktforscher – erfordert hohe kognitive Fähigkeiten von den Probanden	**niedrig** + wesentlich unabhängiger von der Anzahl der Eigenschaften + einfach zu implementieren + geringe Anforderungen an die kognitiven Fähigkeiten der Probanden und das Methodenwissen der Marktforscher
Anzahl Eigenschaften	**wenige (max. 6)** – Gefahr von Vereinfachungsstrategien bei der Beantwortung	**viele** + auch geeignet bei Problemstellungen mit großer Anzahl an Eigenschaften
Anwendbarkeit für Produkte mit hohem Innovationsgrad	**nur bedingt geeignet** – Auswahl der relevanten Eigenschaften problematisch, da diese bei Innovationen meist nicht bekannt – Anwendbarkeit nimmt mit zunehmendem Innovationsgrad ab	**besser geeignet** + Vorauswahl der relevanten Eigenschaften nicht erforderlich, da insgesamt eine größere Anzahl an Eigenschaften betrachtet werden kann
Interviewformen	**persönlich** **schriftlich** **bedingt telefonisch** **online**	**persönlich** **schriftlich** **telefonisch** **online**
Zeitaufwand	**hoch** – höherer Zeitaufwand sowohl für die Konzeption als auch die Datenauswertung der Studie – längere Interviewzeiten	**niedrig** + Datenerhebung als auch Auswertung sind einfach und weniger zeitaufwendig
Kosten	**hoch** – hohe Kosten meist durch persönliche Interviews bedingt	**niedrig** + niedrigere Kosten durch kürzere ggf. telefonische Interviews
Modellannahmen	**flexibel** + Nicht-Linearität in der Teilnutzenfunktion kann berücksichtigt werden	**fest definiert** – additives Teilnutzenmodell wird vorausgesetzt
Berücksichtigung von Interaktionseffekten	**grundsätzlich möglich** + bei einem entsprechenden Experimentdesign ist die Ermittlung von Interaktionseffekten explizit möglich	**nur in Ausnahmefällen** – bei isolierter Beurteilung einzelner Eigenschaften ist kein tatsächlicher Trade-Off erforderlich – Gefahr, daß redundante Eigenschaften durch separate Bewertung doppelt gezählt werden

Quelle: Eigene Erstellung

Ein wesentliches Unterscheidungsmerkmal der beiden Ansätze stellt die als genau entgegengesetzt interpretierbare **Befragungsform** dar. Wie bereits erwähnt, basiert die Conjoint-Analyse als dekompositionelles Verfahren zur Ermittlung der Bedeutungsgewichte von Eigenschaften auf einer indirekten Befragung der Probanden. Die Bewertung ganzheitlicher Produkte oder Alternativen ist der realen Kaufsituation relativ ähnlich und erhöht dadurch den Realitätsbezug der Befragung. Demgegenüber werden bei der Self-Explicated-Methode die Probanden direkt nach der Wichtigkeit der einzelnen Eigenschaften gefragt. Die in der praktischen Anwendung zum Teil unklare Definition von „Wichtigkeit" stellt allerdings ein wesentliches Problem der Methode dar. Je nach Operationalisierung ist für die Probanden teilweise nicht eindeutig, was mit dem Konzept der Wichtigkeit gemeint ist. Eine weitere Gefahr der direkten Befragungsform liegt in möglichen Verzerrungen bei der Schätzung der Teilnutzenfunktionen. Häufig werden wichtige Eigenschaften unterschätzt und die Bedeutung von weniger wichtigen Eigenschaften überschätzt. Dieses Dilemma tritt insbesondere bei sozial erwünschtem Antwortverhalten auf.

Zwar wird mit einer ganzheitlichen Betrachtung von Produkten der Realitätsbezug der Conjoint-Analyse im Vergleich zur Self-Explicated-Methode maßgeblich erhöht, jedoch ist damit auch eine wesentlich höhere **Komplexität** der Aufgabenstellung verbunden. Die Beurteilungsaufgabe stellt hohe Anforderungen an die kognitiven Fähigkeiten der Probanden, da bei der Bewertung der Stimuli viele Informationen gleichzeitig verarbeitet werden müssen. Aber auch für den Marktforscher erfordert die Konzeption einer Conjoint-Analyse umfangreiches Methodenwissen sowohl für das Experimentdesign als auch für die abschließende Daten- und Ergebnisauswertung. Ein wichtiger Vorteil der Self-Explicated-Methode in diesem Zusammenhang ist ihre Einfachheit sowohl in der Implementierung als auch hinsichtlich der Anforderungen an die Probanden. Diese wird auch häufig als entscheidender Vorteil der Methode angeführt. Die Komplexität der Aufgabenstellung bei der Conjoint-Analyse wird zusätzlich von der **Anzahl zu untersuchender Eigenschaften** beeinflußt. Werden die zu beurteilenden Stimuli durch (wesentlich) mehr als 6 Eigenschaften beschrieben, können Validitätsprobleme infolge einer Informationsüberlastung der Probanden auftreten. Um dieser Informationsüberlastung entgegenzutreten, setzen Probanden meist unbewußt Vereinfachungsstrategien ein. Eine Form der Vereinfachung besteht darin, bei der Be-

urteilung der Stimuli lediglich auf eine Teilmenge der abzuwägenden Eigenschaften zu achten. Solche Vereinfachungsstrategien können die Validität der Schätzergebnisse allerdings in hohem Maße beeinträchtigen. Im Gegensatz dazu ist das Self-Explicated-Modell auch für Problemstellungen mit vielen Eigenschaften geeignet, denn der Grad der Komplexität ist insgesamt niedriger und wird weniger stark von der Zahl zu untersuchender Eigenschaften beeinflußt.[117] Teilweise in Zusammenhang mit der Anzahl zu untersuchender Eigenschaften steht die **Anwendbarkeit** der Verfahren **für innovative Produkte**. Voraussetzung jeder Variante der Conjoint-Analyse ist, daß alle relevanten Eigenschaften berücksichtigt werden. Je höher jedoch der Innovationsgrad eines Produkts, desto weniger können diese a priori festgelegt werden. Da die Self-Explicated-Methode eine höhere Anzahl an Eigenschaften zuläßt, ist in ihrer Anwendung keine solche a priori Auswahl erforderlich. Sie ist damit für Produkte mit hohem Innovationsgrad tendenziell besser geeignet.[118]

In Zusammenhang mit der Komplexität der Aufgabe sowohl für Marktforscher als auch für die Probanden steigt der **Zeit-** und damit verbunden der **Kostenaufwand**, den der Einsatz einer Methode verursacht. So werden bei der Conjoint-Analyse häufig persönliche Interviews eingesetzt, was die Kosten der Befragung wesentlich erhöht. Dagegen kann aufgrund der Einfachheit der Abfrage eine direkte Befragung zur Wichtigkeit einzelner Eigenschaften wesentlich schneller und kostengünstiger, beispielsweise im Rahmen eines Telefoninterviews, durchgeführt werden.[119] Die **Interviewform** hat also neben dem erforderlichen Zeitaufwand bei der Datenerhebung und Analyse großen Einfluß auf die entstehenden Kosten einer Studie zur Messung von Nachfragerpräferenzen. Unter diesen

---

117 Einschränkend muß angemerkt werden, daß dieser Vorteil der Self-Explicated-Methode an Gewicht verliert, sofern bei der Conjoint-Analyse ein „Bridging"-Design angewendet wird. In diesem Fall kann auch bei der rein dekompositionellen Conjoint-Analyse eine größere Anzahl an Eigenschaften betrachtet werden.

118 Zur Anwendbarkeit der Verfahren bei Produkten mit hohem Innovationsgrad vgl. Clement (2000), S. 162 ff.

119 Srinivasan/Wyner (1989) schlagen mit CASEMAP eine computergestützte Self-Explicated-Methode vor, die explizit für den Einsatz bei Telefoninterviews konzipiert wurde.

Anwendungsgesichtspunkten überwiegen die Vorteile der Self-Explicated-Methode eindeutig gegenüber der wesentlich komplexeren Conjoint-Analyse.[120]

Aus methodischer Sicht zeichnet sich die Conjoint-Analyse durch eine wesentlich höhere Flexibilität aus. Während bei der Self-Explicated-Methode das additive Teilnutzenmodell als gültig vorausgesetzt wird, sind die Schätzergebnisse einer Conjoint-Analyse auch bei Gültigkeit eines multiplikativen Teilnutzenmodells konsistent. Durch diese Flexibilität der **Modellannahmen** stellen Fehlspezifikationen bei der Conjoint-Analyse ein weniger gravierendes Problem dar als bei den rein kompositionellen Modellen. Ebenfalls flexibler ist die Conjoint-Analyse bei der **Berücksichtigung von Interaktionseffekten**. Während bei entsprechendem Experimentdesign Interaktionseffekte explizit ermittelt werden können, ist dies beim Self-Explicated-Verfahren nur in Ausnahmefällen möglich. Vielmehr wird durch die separate Beurteilung der einzelnen Eigenschaften den Probanden gar kein Trade-Off zwischen den Eigenschaften im eigentlichen Sinn abverlangt.[121] Mit der isolierten Bewertung ist zudem die Gefahr verbunden, daß redundante Eigenschaften doppelt oder noch häufiger gezählt werden und die Bedeutungsgewichte der Eigenschaften dadurch viel höher ausfallen, als sie tatsächlich sind.

Summiert man die angeführten positiven und negativen Auswirkungen für jede der beiden Methoden, zeigt sich tendenziell ein ausgeglichenes Bild. Inhaltliche und methodische Aspekte, so wie der Realitätsbezug der Aufgabenstellung oder die flexibleren Modellannahmen sprechen für den Einsatz der Conjoint-Analyse. Unter Implementierungs- und Anwendungsgesichtspunkten wie z.B. dem not-

---

120 Es sei darauf hingewiesen, daß die Durchführung einer Conjoint-Analyse im Rahmen eines Telefoninterviews nicht prinzipiell ausgeschlossen ist. Sie ist dennoch auf den Einsatz von Paarvergleichen beschränkt, die zudem anhand möglichst weniger Eigenschaften beschrieben sein sollten (vgl. Moore (1992), S. 249). Diese Einschränkungen in der Konzeption des Conjoint-Interviews sind vermutlich entscheidendes Kriterium dafür, daß diese Interviewform in der kommerziellen Anwendung von Conjoint-Analysen eine sehr untergeordnete Rolle spielt (vgl. Wittink/Cattin (1989), S. 92 und Wittink/Vriens/Burhenne (1994), S. 44).

121 Mit dem Vorschlag von Srinivasan (1988), zunächst die wichtigste Eigenschaft zu identifizieren und diese als Ankerpunkt für die Wichtigkeit der anderen Eigenschaften zu verwenden, wird eine relative Bewertung aller Eigenschaften vorgenommen (vgl. Srinivasan (1988), S. 296). Die Eigenschaften werden bei dieser Variante der Self-Explicated-Methode nicht mehr völlig isoliert voneinander betrachtet und die damit verbundenen Nachteile teilweise reduziert.

wendigen Zeitaufwand und entstehenden Kosten müßte dagegen der Self-Explicated-Ansatz favorisiert werden. Zwar wird die Einfachheit der kompositionellen Verfahren in der Literatur durchgängig als Hauptvorteil angeführt, jedoch beschränken sich die Anwendungsempfehlungen meist auf Problemstellungen mit einer großen Anzahl an Eigenschaften. Fast immer wird die Conjoint-Analyse aus theoretischer Sicht der Self-Explicated-Methode zumindest implizit als überlegen angesehen.[122] Begründet wird dies vor allem mit der höheren Prognosegüte der Ergebnisse bedingt durch den stärkeren Realitätsbezug der Beurteilungsaufgabe. Ob sich die diskutierten Probleme der Self-Explicated-Methode in erwarteter Form nachteilig auf die Güte der Ergebnisse auswirken, kann empirisch überprüft werden. Ist dies nicht der Fall, wird das Argument hoher Zeit- und Kostenersparnis sowohl in der Datenerhebung als auch der Datenauswertung von zentraler Bedeutung sein. Einen ersten Schritt in diese Richtung stellt die folgende Zusammenfassung bisher durchgeführter empirischer Vergleiche zwischen der rein dekompositionellen Conjoint-Analyse und der Self-Explicated-Methode dar.

2.3.2 Bisherige empirische Vergleiche

In Anlehnung an Sattler/Hensel-Börner (2000) wird ein umfassender Überblick über bisherige Vergleichsstudien zwischen der Self-Explicated-Methode und der rein dekompositionellen Conjoint-Analyse gegeben. In der folgenden Übersicht werden zusätzlich aktuelle und in der deutschsprachigen Literatur veröffentlichte Studien diskutiert. Eine weitere Ergänzung betrifft eine Aufführung der Studien, welche die Methoden hinsichtlich der internen und diskriminierenden Validität vergleichen. Diese werden im Hinblick auf die im Rahmen der vorliegenden Arbeit durchgeführte empirische Untersuchung einbezogen.[123] Entsprechend der theoretischen Diskussion wird die Conjoint-Analyse insgesamt als dekompositionelles Verfahren gegenüber der Self-Explicated-Methode auf der anderen Seite betrachtet. Die unterschiedlichen Varianten werden aus den bereits oben ange-

---

[122] Dies wird bereits in der Darstellung der Vor- und Nachteile der Self-Explicated-Methode der meisten Autoren deutlich. Während die Einfachheit der Methode als Hauptargument in einem einführenden Satz erwähnt wird, folgt daran eine umfassende Auflistung aller Nachteile der Methode (vgl. hierzu Green/Srinivasan (1990), S. 9 f. und Vriens (1995), S. 7).

[123] Die Abgrenzung der Gütemaße wurde ebenfalls an der im Rahmen dieser Arbeit durchgeführten Untersuchung orientiert. Aus diesem Grund kann es bei der Systematisierung zu Abweichungen im Vergleich zu Sattler/Hensel-Börner (2000) kommen. Zur Abgrenzung der Gütemaße vgl. Kapitel 2.2.2.1.

führten Gründen auch bei der Zusammenfassung der empirischen Ergebnisse nicht explizit differenziert. Die Studien sind daher in der Tabelle nach den eingesetzten Gütemaßen sortiert und innerhalb dieses Kriteriums chronologisch angeordnet.[124] Durch Fettdruck hervorgehoben sind die Ergebnisse, die nach der theoretischen Diskussion erwartet werden, d.h. die Ergebnisse, die hinsichtlich der Prognoseergebnisse eine Überlegenheit der Conjoint-Analyse aufweisen. Des weiteren sind in Tabelle 2-3 der Untersuchungsgegenstand, die Anzahl untersuchter Eigenschaften sowie der Stichprobenumfang angegeben, um wiederum mögliche systematische Einflüsse dieser Untersuchungsmerkmale auf die Ergebnisse identifizieren zu können.

[124] Sofern einzelne Studien Ergebnisse zu verschiedenen Gütemaßen liefern, kann es zu Mehrfachnennung einer Studie in der Tabelle kommen.

Tabelle 2-3: Empirische Vergleiche zwischen der Conjoint-Analyse und der Self-Explicated-Methode (Teil 1)

Studie	**Verglichene Methoden**	**Gütemaße**	**Ergebnisse**	**Produkte**	**Anzahl Eigen-schaften**	**Stich-probe**	**Experi-ment-design**
Heeler/Okechuku/ Reid (1979)	Self-Explicated Conjoint	Reliabilität	Self-Explicated besser als Conjoint a)	Elektrischer Mixer	10	98 b)	between
Leigh/MacKay/ Summers (1984)	Self-Explicated Conjoint	Reliabilität	Self-Explicated sig. besser als Conjoint	Taschen-rechner	5	122 b)	between
Green/Krieger/ Agarwal (1993)	Self-Explicated Conjoint	Reliabilität	Self-Explicated sig. besser als Conjoint	Auto	8	133 b)	within und between
Dorsch/Teas (1992)	Self-Explicated Conjoint	Diskriminierende Validität	Unterschiede bestehen a)	Arbeitsplatz-wahl	9	69 b)	within
Müller-Hagedorn/ Sewing/Toporowski (1993)	Self-Explicated Conjoint	Diskriminierende Validität	sig. Unterschiede	Fernseher Erfrischungs-getränk	5 5	120 50	within within
Stallmeier (1993)	Self-Explicated Conjoint	Diskriminierende Validität	Unterschiede bestehen a)	Kurort-marketing	3 und 5	268	between und within
Baalbaki/Malhotra (1995)	Self-Explicated Conjoint	Diskriminierende Validität	keine Unterschiede	Marketing-Strategien	18	74	within
Green/Carmone/ Wind (1972)	Self-Explicated Conjoint	Interne Validität	keine Unterschiede a)	Kundenkarte	3	43	within
Green/Goldberg/ Montemayor (1981)	Self-Explicated Conjoint	Interne Validität	Self-Explicated besser als Conjoint a)	Antibiotikum	8	404	within

Tabelle 2-3: Empirische Vergleiche zwischen Conjoint-Analyse und Self-Explicated-Methode (Teil 2)

Studie	**Verglichene Methoden**	**Gütemaße**	**Ergebnisse**	**Produkte**	**Anzahl Eigen-schaften**	**Stich-probe**	**Experi-ment-design**
Rao/Solgaard (1978)	Self-Explicated Conjoint	Prognosevalidität	keine Unterschiede a)	Professor	3	29 b)	within
Neslin (1981)	Self-Explicated Conjoint	Prognosevalidität	**Conjoint sig. besser als Self-Explicated**	Arztpraxis	5 (+3)	112	within
Goldberg (1980) (zitiert nach Green/Goldberg/ Montemayor (1981))	Self-Explicated Conjoint	Prognosevalidität	**Conjoint besser als Self-Explicated a)**	Telekom-munikations-zubehör	7	324	within
Green/Goldberg/ Wiley (1982)	Self-Explicated Conjoint	Prognosevalidität	**Conjoint sig. besser als Self-Explicated**	neues Haus-haltsgerät	7	476	within
Cattin/Hermet/Pioche (1982)	Self-Explicated Conjoint	Prognosevalidität	**Conjoint besser als Self-Explicated a)**	Bank Service	5	42	within
Akaah/Korgaonkar (1983)	Self-Explicated Conjoint	Prognosevalidität	**Conjoint sig. besser als Self-Explicated**	HMO-Pläne	6	80	within
Leigh/MacKay/ Summers (1984)	Self-Explicated Conjoint	Prognosevalidität	keine sig. Unterschiede	Taschen-rechner	5	122 b)	between
Green/Helsen (1989)	Self-Explicated Conjoint	Prognosevalidität	keine sig. Unterschiede	Appartment	6	99 b)	within
Agarwal/Green (1991)	Self-Explicated Conjoint	Prognosevalidität	Ergebnisse variieren über Kriterien	Appartment	6	170 b)	within

Tabelle 2-3: Empirische Vergleiche zwischen Conjoint-Analyse und Self-Explicated-Methode (Teil 3)

Studie	**Verglichene Methoden**	**Gütemaße**	**Ergebnisse**	**Produkte**	**Anzahl Eigenschaften**	**Stichprobe**	**Experimentdesign**
van der Lans/ Heiser (1992)	Self-Explicated Conjoint	Prognosevalidität	Ergebnisse variieren über Kriterien	Appartment	6	177 b)	within
Green/Krieger/ Agarwal (1993)	Self-Explicated Conjoint	Prognosevalidität	Ergebnisse variieren über Kriterien	Auto	8	133 b)	within und between
Huber et al. (1993)	Self-Explicated Conjoint	Prognosevalidität	keine sig. Unterschiede	Kühlschrank	5 bzw. 9	393	within
Pullman/Dodson/ Moore (1999)	Self-Explicated Conjoint	Prognosevalidität	**Conjoint sig. besser als Self-Explicated**	Nationalpark	17	390 b)	between
Huber/Danesghar/ Ford (1971)	Self-Explicated Conjoint	Externe Validität	Self-Explicated besser als Conjoint a)	Arbeitsplatzwahl	5	30 b)	within
Wright/Kriewall (1980)	Self-Explicated Conjoint	Externe Validität	Self-Explicated sig. besser als Conjoint	Studienortwahl	5	120	within
Srinivasan (1988)	Self-Explicated Conjoint c)	Externe Validität	keine sig. Unterschiede	Arbeitsplatzwahl	8	45/48 b) d)	between

a) keine Aussage zur Signifikanz
b) Befragt wurden Studierende
c) Einzige Ausnahme, bei der die Trade-Off Conjoint Analysis mit der Self-Explicated-Methode verglichen wurde
d) Stichprobenumfang variiert, da Vergleich über 2 verschiedene Untersuchungen gemacht wird

Quelle: Eigene Erstellung in teilweiser Anlehnung an Sattler/Hensel-Börner (2000), S. 128 f.

Vergleiche hinsichtlich der **diskriminierenden Validität** zeigen in 3 von 4 Fällen, daß je nach Methode unterschiedliche Präferenzstrukturen ermittelt werden. Dieses Ergebnis impliziert, daß die Art der Befragung die ermittelten Präferenzstrukturen beeinflußt. Ob die Ergebnisse der Conjoint-Analyse wie erwartet die valideren sind und den tatsächlichen Präferenzstrukturen der Probanden näher kommen, kann anhand dieses Kriteriums nicht beurteilt werden. Die diskriminierende Validität läßt lediglich eine Aussage zu Unterschieden zwischen den Ergebnissen zu. Von größerer Bedeutung sind daher die Vergleiche hinsichtlich der anderen Gütekriterien.

Für jede Datenerhebungsmethode kann mit der **Reliabilität** die Konsistenz des Antwortverhaltens der Probanden geprüft werden. Die Ergebnisse sind diesbezüglich eindeutig. Bei allen 3 vorliegenden Vergleichen führt die Self-Explicated-Methode zu reliableren Ergebnissen. Dies stärkt das Argument, daß ein höherer Grad an Komplexität konsistentes Verhalten der Probanden erschwert. Bei einer einfacheren Aufgabe wie der Beurteilung einzelner Eigenschaften fällt es den Probanden dagegen leichter, konsistent zu antworten. Deutlich wird zudem, daß dieses Ergebnis von der Anzahl untersuchter Eigenschaften nicht beeinflußt wird. Nicht nur bei relativ vielen Eigenschaften (8 bzw. 10) ist die Self-Explicated-Methode der Conjoint-Analyse überlegen, sondern bereits bei nur 5 Eigenschaften schneidet letztere deutlich schlechter ab.[125]

Im Gegensatz zur Reliabilität testet die **interne Validität**, inwieweit das Antwortverhalten der Probanden durch das Modell erklärt werden kann. Sie gibt an, wie gut sich die empirisch geschätzten Präferenzen an die tatsächlich beobachteten Präferenzen anpassen.[126] Diesbezüglich weisen bisherige empirische Befunde darauf hin, daß die Self-Explicated-Methode mindestens zu genauso guten Ergebnissen führt wie die rein dekompositionelle Conjoint-Analyse. Berücksichtigt man bei der Interpretation dieser Ergebnisse, daß für die Self-Explicated-Methode dieser Vergleich über einen Test der internen Validität im engeren Sin-

---

125 Auf eine Hervorhebung dieser erwarteten Ergebnisse in Tabelle 2-3 wurde verzichtet, da lediglich die Befunde hinsichtlich der Prognosegüte verdeutlicht werden sollen.

126 Es sei an dieser Stelle angemerkt, daß ein Vergleich zwischen der Conjoint-Analyse mit der Self-Explicated-Methode, zumindest für letztgenannte, sogar über eine Prüfung der internen Validität im engeren Sinne hinausgeht. Für die Self-Explicated-Methode stellt dies bereits eine Prüfung der Prognosevalidität dar, da in den betrachteten Studien jeweils die beobachteten Präferenzen der Conjoint-Aufgabe prognostiziert wurden.

ne hinausgeht, gewinnt das gute Abschneiden des Verfahrens zusätzlich an Bedeutung.

Von besonderem Interesse für den Einsatz einer Methode zur Präferenzmessung ist deren Eignung, Präferenzen für alternative, vor allem neue Produkte prognostizieren zu können. Auf Basis der theoretischen Argumentation sprechen die Vorteile überwiegend für die dekompositionelle Vorgehensweise. Analysiert man diesbezüglich bisherige Forschungsergebnisse, zeigt sich dagegen keine eindeutige Überlegenheit der Conjoint-Analyse. Während sich im Hinblick auf die Prognosegüte für knapp die Hälfte der Vergleiche eine (z.T. signifikant) höhere **Prognosevalidität** für die Conjoint-Analyse ergab, kann in entsprechend vielen Fällen kein Unterschied zwischen den Verfahren nachgewiesen oder keine eindeutige Aussage gemacht werden. Prüft man die Prognosegüte anhand der **externen Validität**, also bezüglich tatsächlicher Entscheidungen, die unabhängig von der Befragung getroffen wurden, schneidet die Conjoint-Analyse deutlich schlechter ab als die Self-Explicated-Methode. In allen 3 Studien, die dieses Gütemaß zur Beurteilung heranziehen, ist die Self-Explicated-Methode gleich gut oder der Conjoint-Analyse sogar überlegen.

Diese Ergebnisse können insgesamt nur unzureichend durch die Besonderheiten der einzelnen Untersuchungen begründet werden. Weder in bezug auf das Untersuchungsobjekt noch auf die Stichprobe kann ein systematischer Einfluß abgeleitet werden. Gleiches gilt für die Anzahl berücksichtigter Eigenschaften. Selbst bei gleicher Anzahl an Eigenschaften widersprechen sich die Befunde. Akaah/Korgaonkar (1983), Green/Helson (1989) sowie Agarwal/Green (1991) betrachten jeweils 6 Eigenschaften und berichten widersprüchliche Ergebnisse. Außerdem können Pullman/Dodson/Moore (1999) zeigen, daß bei Berücksichtigung von 17 Eigenschaften die Conjoint-Analyse zu signifikant höherer Prognosevalidität führt als die Self-Explicated-Methode. Umgekehrt schneidet sie in anderen Studien bereits bei 5 Eigenschaften schlechter ab als die Self-Explicated-Methode.[127] Zudem sei angemerkt, daß die Self-Explicated-Methode und die

[127] Vgl. hierzu die Studien von Huber/Danesghar/Ford (1971), Srinivasan (1988), Leigh/MacKay/Summers (1984).

Conjoint-Analyse auf Basis der angeführten Argumente ganzheitlich betrachtet wurden. Eine Aussage über systematische Einflüsse der eingesetzten Variante der beiden Verfahren ist dadurch nicht möglich. Da bisher keine eindeutigen (signifikanten) Unterschiede zwischen den einzelnen Varianten nachgewiesen werden konnten (vgl. Abschnitt 2.1 und 2.2), liegt die Vermutung nahe, daß auch dieser Einflußfaktor keine Erklärung für die vorliegenden Ergebnisse liefert. Das Gesamtergebnis muß dennoch vor dem Hintergrund eines möglichen Informationsverlustes durch die zusammengefaßte Betrachtung interpretiert werden.

Insgesamt lassen 20 Befunde einen direkten Vergleich der beiden Ansätze zur Präferenzmessung hinsichtlich verschiedener Gütemaße zu.[128] Bei diesen 20 Vergleichen kann nur in 6 (30%) Fällen eine (nur teilweise signifikante) Überlegenheit der Conjoint-Analyse nachgewiesen werden. Fast genauso oft (25% der Fälle) schneidet die Self-Explicated-Methode besser ab und weitere 8 Studien zeigen keine Unterschiede in der Ergebnisgüte bzw. lassen keine eindeutige Aussage zu. Bereits vor dem Hintergrund dieser Ergebnisse muß der dominante Einsatz der rein dekompositionellen Conjoint-Analyse insbesondere in der praktischen Anwendung kritisch hinterfragt werden. Auch wenn zahlreiche Argumente und Vorteile für das dekompositionelle Vorgehen sprechen, kann diese Überlegenheit insbesondere bei der Prognose von tatsächlichem Kaufverhalten empirisch bisher nicht belegt werden. Vielmehr führen die wesentlich einfacheren Self-Explicated-Verfahren zu vergleichbaren oder teilweise besseren Ergebnissen und stellen vor allem unter dem Aspekt der Implementierung und unter Zeit- und Kostengesichtspunkten in der Durchführung durchaus eine Alternative dar. Offen ist an dieser Stelle noch, ob diese Aussage auch im Vergleich zu hybriden, insbesondere den neueren computergestützten Conjoint-Varianten bestehen bleiben kann. Eine entsprechende Prüfung wird nach der Darstellung hybrider Verfahren durchgeführt.

---

[128] Nicht berücksichtigt in dieser Zusammenfassung sind die Vergleiche hinsichtlich der diskriminierenden Validität, da diese keinen direkten Vergleich zugunsten einer der beiden Methoden zulassen.

2.4 Hybride Verfahren

Entsprechend den Ausführungen zu den kompositionellen sowie dekompositionellen Verfahren sollen an dieser Stelle die hybriden Ansätze zur Präferenzmessung in erster Linie systematisiert werden.[129] Schwerpunkt der Systematisierung ist, die verschiedenen Varianten entsprechend ihrer konzeptionellen Besonderheiten voneinander abzugrenzen. Allen hybriden Verfahren ist gemein, daß die Datenerhebung sowohl einen kompositionellen (Self-Explicated-Phase) als auch einen dekompositionellen (Conjoint-Phase) Interviewteil umfaßt. Die beiden Vorgehensweisen werden mit dem Ziel kombiniert, die jeweiligen Vorteile der einzelnen Schritte auszunutzen.[130] Welche Variante der Self-Explicated-Methode bzw. der Conjoint-Analyse zum Einsatz kommt, ist unterschiedlich. Von größerer Bedeutung für die Abgrenzung der hybriden Ansätze ist, wie diese beiden Verfahrensschritte aufeinander abgestimmt werden. Dabei kann der Conjoint-Teil unabhängig von der Self-Explicated-Befragung und damit a priori konzipiert werden. Adaptive hybride Conjoint-Analysen dagegen setzen die aus der Self-Explicated-Phase gewonnenen Informationen ein, um den Conjoint-Teil der Befragung individuell auf den Probanden abzustimmen. Unter dem Begriff "adaptiv" werden im Rahmen dieser Arbeit alle Verfahren subsumiert, bei denen die Gestaltung eines zweiten Befragungsteils an die Antworten aus einer ersten Interviewphase angepaßt wird. Diese Begriffsabgrenzung wird dabei unabhängig vom Ausmaß der Anpassungen vorgenommen. Abhängig von der eingesetzten Conjoint-Variante kann diese Anpassung unterschiedlich stark ausgeprägt sein.[131] Mit dieser Begriffsabgrenzung werden somit auch individualisierte bzw. „customized" Ansätze als adaptive Verfahren betrachtet.[132] Abbildung 2-4 gibt einen Überblick über hybride Ansätze zur Präferenzmessung.

---

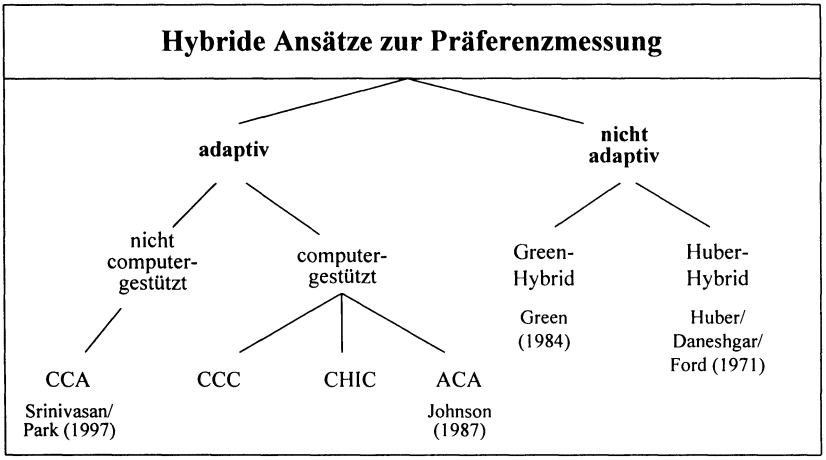
129 Die im Rahmen dieser Arbeit untersuchten Verfahren werden in Kapitel 3 ausführlich dargestellt.

130 Die Vor- und Nachteile der beiden unterschiedlichen Vorgehensweisen wurden in Kapitel 2.3.1 diskutiert.

131 Vgl. hierzu Kapitel 3.

132 Es sei zudem angemerkt, daß auch in der bisherigen Literatur diese Begriffe nicht eindeutig voneinander abgegrenzt sind. So wird beispielsweise die Vorgehensweise der Adaptive Conjoint Analysis (ACA) im Handbuch zur Software mit „...customizing the interviews for each respondent ..." beschrieben (vgl. Metegrano (1994), S. 1-3).

Abbildung 2-4: Systematisierung hybrider Ansätze



Quelle: Eigene Erstellung

2.4.1 Nicht adaptive hybride Verfahren

Ziel einer hybriden Methode ist es, auch bei Einsatz einer dekompositionellen Vorgehensweise, eine möglichst große Anzahl an Eigenschaften in einer Untersuchung berücksichtigen zu können.[133] Während die traditionelle Voll-Profil-Methode auf max. 6 Eigenschaften beschränkt ist,[134] sind bei kommerziellen Anwendungen in den meisten Fällen wesentlich mehr Eigenschaften untersuchungsrelevant.[135] Grundgedanke der traditionellen hybriden Vorgehensweise ist daher, die zu beurteilenden Stimuli der Conjoint-Phase auf mehrere Probanden aufzuteilen und dadurch einer Überforderung der Probanden vorzubeugen.[136] Während jede Testperson in einer Self-Explicated-Phase alle Eigenschaften bewertet, wird ihr im Conjoint-Teil lediglich eine Teilmenge aller kombinierten Stimuli vorgelegt. Die Stimuli werden dabei einem sog. Masterdesign entnommen und in einem ausgeglichenen Verhältnis auf die verschiedenen Probanden aufgeteilt. Die Datenauswertung des Verfahrens basiert dann auf dem Prinzip, Probanden mit einer ähnlichen Nutzenstruktur (ermittelt aus der Self-Explicated-Phase) zu ho-

---

133 Einen umfassenden Überblick zur Vorgehensweise hybrider Conjoint-Analysen gibt Green (1984).

134 Vgl. Green/Srinivasan (1990), S. 8.

135 Vgl. Green (1984), S. 156.

136 Das Hauptproblem einer großen Anzahl an Eigenschaften liegt darin, daß für die stabile Schätzung der Teilnutzenwerte sehr viele Stimuli konstruiert und damit auch bewertet werden müssen.

mogenen Gruppen zusammenzufassen und die Conjoint-Daten segmentweise auszuwerten. Die individuellen Nutzenfunktionen auf Basis der Self-Explicated-Befragung werden auf diese Weise um einen segmentspezifischen Modellteil erweitert.

Neben dieser grundsätzlichen Konzeption der hybriden Verfahren treten je nach Autor Modifikationen dahingehend auf, welche Varianten der Self-Explicated-Methode oder Conjoint-Analyse eingesetzt werden. Die Gestaltungsspielräume seien an folgenden Beispielen verdeutlicht. Während bei Green/Goldberg/Montemayer (1981) in der Self-Explicated-Phase sowohl die Eigenschaftsausprägungen beurteilt als auch die Wichtigkeit der Eigenschaften erhoben werden (gewichtete Self-Explicated-Methode),[137] umfaßt der Self-Explicated-Teil der Huber-Hybrid-Variante ausschließlich eine Bewertung der Ausprägungen. Die Wichtigkeit der Eigenschaften wird anschließend durch Regression bestimmt, wobei die Bewertungen der Ausprägungen als unabhängige und die Bewertungen für die Gesamtkonzepte der Conjoint-Phase als abhängige Variable in die Schätzung eingehen.[138] Während in der Conjoint-Phase in den meisten Studien bisher der Einsatz einer Ranking- oder Rating-Aufgabe überwiegt, ist grundsätzlich auch hier ein alternatives Vorgehen denkbar. Moore/Gray-Lee/Louviere (1998) beispielsweise verbinden echte Auswahlentscheidungen einer Choice-Based Conjoint-Analyse mit zuvor erhobenen Self-Explicated-Daten und sprechen von einem hybriden Choice-Modell.[139]

Mit der Aufteilung aller Stimuli auf verschiedene Probanden geht allerdings die Möglichkeit einer individuellen Schätzung der Teilnutzenwerte *allein auf Basis der Conjoint-Daten* verloren. Für jeden Probanden werden lediglich Präferenzen gegenüber einem Teil aller zu bewertenden Stimuli erhoben. Dies liefert keine ausreichende Datenmenge, um individuelle Nutzenfunktionen schätzen zu können. Um dieses Defizit der traditionellen hybriden Verfahren zu beheben, wurden von Green/Krieger (1996) und Baier/Säuberlich (1997) neuere Schätzmethoden vorgeschlagen, die eine vollständig individualisierte Auswertung hybrider Con-

---

137 Vgl. Green/Goldberg/Montemayer (1981), S. 34.

138 Vgl. Hoepfl/Huber (1970), S. 411; Huber/Daneshgar/Ford (1971), S. 271. Zur Abgrenzung zwischen Huber- und Green-Hybrid vgl. auch Akaah/Korgaonkar (1983), S. 187 f.

139 Vgl. Moore/Gray-Lee/Louviere (1998), S. 197.

joint-Analysen ermöglichen. Dabei werden sowohl die Self-Explicated-Daten als auch die Conjoint-Daten auf individuellem Niveau ausgewertet.[140]

2.4.2 Adaptive hybride Ansätze

2.4.2.1 Nicht computergestützte Verfahren

Sowohl die Handhabung vieler Eigenschaften als auch eine individuelle Schätzung der Teilnutzenwerte separat für die Self-Explicated-Phase und die Conjoint-Phase ermöglichen die adaptiven hybriden Conjoint-Analysen. Der Ansatzpunkt zur Berücksichtigung vieler Eigenschaften liegt nicht in der Reduzierung der Stimuli für jede Person, sondern hauptsächlich in der Identifikation der für den Conjoint-Teil tatsächlich relevanten Eigenschaften. Statt in der Self-Explicated- und der Conjoint-Phase jeweils alle Eigenschaften beurteilen zu lassen, werden in der letzteren nur die für die Auskunftsperson wichtigsten Eigenschaften integriert.[141] Srinivasan/Park (1997) schlagen mit der **Customized Conjoint Analysis (CCA)** eine Methode vor, bei der in einem Self-Explicated-Interview alle Eigenschaften bewertet werden. Aufbauend auf den Antworten in diesem ersten Interviewteil werden anschließend für jeden Probanden die individuell wichtigsten Eigenschaften identifiziert und in einer zweiten Befragung - dem Conjoint-Interview - ein zweites Mal beurteilt. Für beide Befragungsschritte werden Nutzenfunktionen für die Eigenschaften auf individuellem Niveau ermittelt und gemäß einer Optimalitätsbedingung anschließend mit denen der Self-Explicated-Phase kombiniert. Da es sich bei der CCA um eine schriftliche Befragung handelt, ist zwischen den beiden Befragungen eine Zeitverzögerung erforderlich, um die Stimuli gemäß den Antworten aus dem Self-Explicated-Teil individuell zu konstruieren. Eine solche wiederholte Befragung der Probanden erhöht allerdings den Erhebungsaufwand und die damit verbundenen Kosten.[142] Zudem kann nicht ausgeschlossen werden, daß sich die Präferenzen der Probanden über die Zeit verändert haben.

---

140 Eine ausführliche Darstellung der Schätzmodelle findet sich bei Green/Krieger (1996), S. 854 ff. und bei Baier/Säuberlich (1997), S. 955 ff.

141 Dieses Vorgehen ist ähnlich der von Schweikl (1985) vorgeschlagenen Idee zur Individualisierung der Conjoint-Analyse. Der wesentliche Unterschied besteht darin, daß es sich bei Schweikl nicht um ein hybrides Verfahren handelt, sondern lediglich nicht relevante Eigenschaften aus der weiteren Betrachtung ausgeschlossen werden. Informationen zu deren Bedeutung werden nicht mehr erhoben.

142 Vgl. Srinivasan/Park (1997), S. 287.

2.4.2.2 Computergestützte Verfahren

Eine unmittelbare Anpassung des Conjoint-Teils an vorangegangene Interviewschritte ist mit Hilfe eines **computergestützten Verfahrens** möglich. Die bekannteste und am weitesten verbreitete Methode ist die **Adaptive Conjoint Analysis (ACA)** von Johnson (1987). In einem vollständig computergestützten Interview werden nach einem Self-Explicated-Teil im dekompositionellen Teil Paarvergleiche (Graded Paired Comparisons) abgefragt, deren Konstruktion jeweils an vorangegangene Antworten angepaßt wird. Die ACA ist bisher die einzige kommerzielle Software für adaptive hybride Verfahren.[143] Zweifelsohne hat sie sich weitestgehend zum Standardinstrument bei der Messung von Präferenzen etabliert. Gleichwohl wird der Einsatz von Paarvergleichen zum einen aufgrund der Ineffizienz in der Datenerhebung häufig kritisiert und spielt bei der Gestaltung rein dekompositioneller Conjoint-Analysen eine sehr untergeordnete Rolle.[144]

Vor diesem Hintergrund werden mit der vorliegenden Arbeit neue adaptive Verfahren vorgeschlagen und der Bereich computergestützter hybrider Verfahren erweitert. Während die neuen Verfahren in Kapitel 3 ausführlich dargestellt werden, soll an dieser Stelle bereits deren Einordnung und Zusammenhang mit bereits bekannten Ansätzen vorgenommen bzw. verdeutlicht werden. Mit der **Customized Computerized Conjoint Analysis (CCC)** wird ein in Anlehnung an die CCA konzipiertes Verfahren vorgestellt, bei der Stimuli in der Conjoint-Phase alle simultan zu bewerten sind. Bei der **Choice-Oriented Individualized Conjoint Analysis (CHIC)** werden den Probanden jeweils Teilmengen aller zu beurteilenden Stimuli gleichzeitig präsentiert. Als eine Aufgabe wird den Probanden eine Auswahlentscheidung abverlangt, um analog zur Choice-Based Conjoint Analysis die Beurteilungsaufgabe so nah wie möglich an der tatsächlichen Kaufsituation zu orientieren. Mit dieser Erweiterung werden im Bereich computergestützter hybrider Verfahren alle drei vorgestellten Gruppen dekompo-

---

143 Carmone/Schaffer (1995), S. 117.

144 Wittink/Cattin (1989) berichten, daß nur in 10% aller kommerziellen Anwendungen in den USA Paarvergleiche eingesetzt werden. (vgl. Wittink/Cattin (1989), S. 93). In einer entsprechenden Untersuchung für Europa wird dieser Trend noch deutlicher. Lediglich in 4% bzw. 5% aller kommerziellen Conjoint-Studien kommen Paarvergleiche (abgesehen von der ACA) zum Einsatz (vgl. Wittink/Vriens/Burhenne (1994), S. 44).

sitioneller Varianten („alle Stimuli", „Teilmenge an Stimuli", „zwei Stimuli") berücksichtigt:

Abbildung 2-5: Zusammenhang zwischen den dekompositionellen und computergestützten hybriden Verfahren

Anzahl simultan zu beurteilende Stimuli	**Dekompositionelle Verfahren**	**Computergestützte hybride Verfahren**
Alle	•traditionelle Conjoint-Analyse mit Ranking oder Rating-Aufgabe	Customized Computerized Conjoint Analysis (CCC)
Teilmenge	•Choice-Based Conjoint Analysis	Choice-Oriented Individualized Conjoint Analysis (CHIC)
Zwei	•Paarvergleiche •Graded Paired Comparison	Adaptive Conjoint Analysis (ACA)

Quelle: Eigene Erstellung

Empirische Vergleiche innerhalb des hybriden Verfahrenszweiges wurden bisher relativ wenige durchgeführt.[145] So liefern Akaah/Korgaonkar (1983) den bisher einzigen expliziten Vergleich innerhalb nicht adaptiver Verfahren zwischen Green- und Huber-Hybrid. Als Ergebnis berichten die Autoren, daß die Unterschiede hinsichtlich der Prognosegüte sehr gering sind und zudem je nach Modellstruktur unterschiedlich ausfallen.[146] Ebenso wird die ACA bisher erst einmal mit einer nicht adaptiven hybriden Variante in einer Studie von Pullman/Dodson/Moore/(1999) verglichen. Die Ergebnisse dieser Untersuchung zeigen, daß keine signifikanten Unterschiede zwischen der ACA und einem traditionellem hybriden Verfahren auftreten bzw. unter Berücksichtigung von „constraints" bei der Schätzung des hybriden Modells, dieses zu einer signifikant höheren Prognosevalidität führt als die ACA.[147] Für die empirische Studie der vorliegenden Arbeit wurde ein umfangreicher Pretest durchgeführt, der gleichzeitig einen ersten empirischen Vergleich zwischen den computergestützten hybriden

[145] Es sei darauf hingewiesen, daß Vergleiche zwischen alternativen Auswertungsmethoden wie sie z.B. von Green/Krieger (1996) durchgeführt werden, nicht Gegenstand der Betrachtung sind, da Schwerpunkt der vorliegenden Arbeit die Datenerhebungsmethode ist.

[146] Vgl. Akaah/Korgaonkar (1983), S. 194.

[147] Vgl. Pullman/Dodson/Moore (1999), S. 131.

Verfahren CCC und ACA liefert. Dabei schneidet die CCC hinsichtlich der Prognosevalidität (teilweise signifikant) besser ab als die ACA.[148]

Die insgesamt geringe Anzahl empirischer Vergleiche läßt sich zumindest teilweise damit begründen, daß eine Gegenüberstellung der hybriden Verfahren mit der rein kompositionellen und rein dekompositionellen Vorgehensweise von größerem Interesse ist. Im folgenden Abschnitt 2.5 wird eine entsprechende Analyse vorgenommen.

2.5 Vergleich hybrider Verfahren mit anderen Ansätzen zur Präferenzmessung

Wie bereits erläutert, werden bei den hybriden Verfahren ein Self-Explicated-Teil und ein Conjoint-Teil mit dem Ziel kombiniert, die Vorteile beider Verfahren auszunutzen. Eine Diskussion der theoretischen Vor- und Nachteile anhand der in Kapitel 2.3.1 definierten Kriterien führt zu dem Ergebnis, daß die hybriden Verfahren hinsichtlich ihrer Ausprägung der verschiedenen Kriterien jeweils zwischen den beiden gegensätzlichen Ansätzen liegen. So können beispielsweise bei einem hybriden Verfahren mehr Eigenschaften betrachtet werden als bei der rein dekompositionellen Conjoint-Analyse, aber nach wie vor weniger als bei der Self-Explicated-Methode. Auf eine erneute Diskussion der theoretischen Vor- und Nachteile der verschiedenen Ansätze wird daher an dieser Stelle verzichtet und der Schwerpunkt auf die Auswertung bisheriger empirischer Ergebnisse gelegt. Diese können zeigen, ob es durch die Kombination der beiden Befragungsteile tatsächlich gelingt, die Schwachstellen der beiden Methoden für sich zu überwinden und die jeweiligen Vorteile auszunutzen. Ist dies der Fall, müßte sich eine Überlegenheit der hybriden Verfahren im Vergleich sowohl zur rein dekompositionellen Conjoint-Analyse als auch der Self-Explicated-Methode hinsichtlich der Güte der Ergebnisse zeigen.

2.5.1 Bisherige empirische Vergleiche zwischen hybriden Verfahren und der rein dekompositionellen Conjoint-Analyse

Entsprechend der Vorgehensweise in den vorangestellten Kapiteln werden die Befunde wieder nach den jeweils verglichenen Methoden angeordnet. Auch hier wird die Systematisierung der Verfahren beibehalten und Tabelle 2-4 entspre-

---

[148] Vgl. Hensel-Börner/Sattler (2000), S. 15 ff.

chend unterteilt. Sofern in einer Studie mehrere Methoden verglichen wurden, kann es zu Mehrfachnennungen dieser Studie kommen, da die vorgenommenen Vergleiche immer paarweise betrachtet werden. Hervorgehoben sind wieder die Ergebnisse, die auf Basis der Konzeption der hybriden Verfahren mit den theoretischen Erwartungen konform gehen. Dementsprechend sind die Ergebnisse fett gedruckt, bei denen mit den hybriden Verfahren eine höhere Validität gegenüber der rein dekompositionellen Conjoint-Analyse erzielt wurde. Ein solches Ergebnis kann gerade deshalb vermutet werden, weil mit der hybriden Vorgehensweise vor allem die Komplexität im Conjoint-Teil reduziert wird. Gleichzeitig wird ein höherer Realitätsbezug als bei der Self-Explicated-Methode allein sichergestellt. Da die mit der Komplexität der Aufgabenstellung bei der Conjoint-Analyse verbundenen Probleme vor allem bei vielen Eigenschaften zum Tragen kommen, müßte sich gerade in diesen Fällen eine Überlegenheit der hybriden Verfahren zeigen.

Tabelle 2-4: Empirische Vergleiche zwischen hybriden Verfahren und der Conjoint-Analyse (Teil 1)

Studie	**Verglichene Methoden**	**Gütemaße**	**Ergebnisse**	**Produkte**	**Anzahl Eigenschaften**	**Stichprobe**	**Experimentdesign**
Traditionell Hybrid vs. Conjoint							
Green/Goldberg/ Montemayor (1981)	Hybrid Conjoint	Interne Validität	**Hybrid sig. besser als Conjoint**	Antibiotikum	8	404	within
Goldberg (1980) (zitiert nach Green/Goldberg/ Montemayer (1981))	Hybrid Conjoint	Prognosevalidität	keine Unterschiede a)	Telekommunikationszubehör	7	324	within
Cattin/Hermet/ Pioche (1982)	Hybrid Conjoint	Prognosevalidität	Conjoint besser als Hybrid a)	Bankservice	5	42	within
Green/Goldberg/ Wiley (1982)	Hybrid Conjoint	Prognosevalidität	keine sig. Unterschiede	neues Haushaltsgerät	7	476	within
Akaah/Korgaonkar (1983)	Hybrid Conjoint	Prognosevalidität	keine sig. Unterschiede	HMO Pläne	6	80	within
Moore/Gray-Lee/ Louviere (1998)	Hybrid Conjoint	Prognosevalidität	**Hybrid sig. besser als Conjoint**	Zahncreme	5	184 b)	within
Pullman/Dodson/ Moore (1999)	Hybrid Conjoint	Prognosevalidität	Conjoint sig. besser als Hybrid	Nationalpark	17	390 b)	between
Huber/Danesghar/ Ford (1971)	Hybrid Conjoint	Externe Validität	Ergebnisse variieren über Kriterien	Arbeitsplatzwahl	5	30 b)	within

Tabelle 2-4: Empirische Vergleiche zwischen hybriden Verfahren und der Conjoint-Analyse (Teil 2)

Studie	Verglichene Methoden	Gütemaße	Ergebnisse	Produkte	Anzahl Eigen-schaften	Stich-probe	Experi-ment-design
ACA vs. Conjoint							
Green/Krieger/Agarwal (1993)	ACA Conjoint	Reliabilität	**ACA sig. besser als Conjoint**	Auto	8	133 b)	between
Agarwal/Green (1991)	ACA Conjoint	Prognosevalidität	Conjoint sig. besser als ACA	Appartment	6	170 b)	within
Mehta/Moore/Pavia (1992)	ACA Conjoint	Prognosevalidität	Ergebnisse variieren über Kriterien	Appartment	4	75 b)	within
Green/Krieger/Agarwal (1993)	ACA Conjoint	Prognosevalidität	Conjoint sig. besser als ACA	Auto	8	133 b)	between
Huber et al. (1993)	ACA Conjoint	Prognosevalidität	**ACA sig. besser als Conjoint**	Kühlschrank	5 bzw. 9	393	within
Orme/Alpert/ Christinsen (1997)	ACA Conjoint	Prognosevalidität	Conjoint sig. besser als ACA	Computer	9	80 b)	within
Pullman/Dodson/ Moore (1999)	ACA Conjoint	Prognosevalidität	Conjoint sig. besser als ACA	Nationalpark	17	390 b)	between

a) keine Aussage zur Signifikanz
b) Befragt wurden Studierende
Quelle: Eigene Erstellung in teilweiser Anlehnung an Sattler/Hensel-Börner (2000), S. 128 f.

Faßt man zunächst die Ergebnisse bzgl. aller Gütemaße bei den Vergleichen der Conjoint-Analyse mit den traditionellen hybriden Verfahren zusammen, weisen letztere nur in 2 der 8 Studien eine höhere Validität auf. Bei den übrigen Vergleichen treten keine Unterschiede auf bzw. variieren die Ergebnisse über die Kriterien. Teilweise schneidet die rein dekompositionelle Conjoint-Analyse sogar besser ab. Dabei wurde in der Mehrzahl der Studien mit einem within-subject-design gearbeitet. In einigen Fällen wurde sogar für die reinen Conjoint-Daten lediglich der zweite Befragungsteil der hybriden Datenerhebung ausgewertet. Die Ergebnisse deuten folglich darauf hin, daß die zusätzlich gewonnenen Informationen durch den Self-Explicated-Teil kaum eine Verbesserung in der Prognosegüte bewirken. Vielmehr können durch Auswertung nur eines Teils der gewonnenen Daten vergleichbare Prognoseergebnisse erzielt werden.

Diese Tendenz zeigt sich in ähnlicher Form bei einem Vergleich der Conjoint-Analyse mit der ACA. Bei 4 von 7 direkten Vergleichen führt die Conjoint-Analyse zu einer signifikant höheren Prognosevalidität. In einem weiteren Fall kann keine eindeutige Aussage gemacht werden und nur zweimal ist die ACA der rein dekompositionellen Conjoint-Analyse überlegen. Ob die höhere Reliabilität der ACA in der Studie von Green/Krieger/Agarwal (1993) durch den Self-Explicated-Teil zurückzuführen ist, kann nur vermutet werden, würde aber den bisherigen Ergebnissen entsprechen. Die Self-Explicated-Verfahren wiesen in den bereits ausgewerteten Studien durchweg eine höhere Reliabilität auf als die Conjoint-Analyse.

Zusammenfassend deutet eine Gegenüberstellung der hybriden Verfahren mit der rein dekompositionellen Conjoint-Analyse darauf hin, daß die Erweiterung des Ansatzes um einen kompositionellen Befragungsteil nicht die erwartete Verbesserung sichert. Vielmehr wird deutlich, daß mit einer dekompositionellen Befragung durchaus vergleichbare, teilweise sogar bessere Ergebnisse im Hinblick auf die Prognosevalidität erzielt werden können. Dies trifft selbst dann zu, wenn in

der Studie eine größere Anzahl von Eigenschaften untersucht werden.[149] Ob auch der Self-Explicated-Teil einer hybriden Variante für sich zu vergleichbar guten Ergebnissen führt, wird im folgenden Kapitel durch Auswertung der entsprechenden empirischen Vergleiche analysiert.

2.5.2 Bisherige empirische Vergleiche zwischen hybriden Verfahren und der Self-Explicated-Methode

Auch bei Vergleichen zwischen den hybriden Methoden mit der Self-Explicated-Methode zeigen sich Parallelen zu den eben diskutierten Befunden. Die Erwartung, durch die Kombination eines kompositionellen mit einem dekompositionellen Befragungsteil bessere Prognoseergebnisse zu erzielen, kann nicht bestätigt werden (vgl. Tabelle 2-5). Nicht einmal eine individuelle Anpassung des Conjoint-Teils führt zu deutlich besseren Untersuchungsergebnissen hinsichtlich der Validität. Vielmehr zeigt sich auch hier die Tendenz, daß bereits durch einen Teil der Informationen entsprechend gute Prognoseergebnisse erzielt werden können. Dieses Argument wird insbesondere dadurch gestärkt, daß nahezu alle hier angeführten Vergleiche auf einem „within-subject-design" basieren. Für den Vergleich der Methoden wurden also keine zusätzlichen Daten erhoben, sondern bereits nach dem ersten Interviewteil separate Schätzungen der Teilnutzenwerte vorgenommen.

---

[149] Vgl. hierzu die Ergebnisse von Green/Krieger/Agarwal (1991), Huber et al. (1993) oder Pullman/Dodson/Moore (1999). Bei diesen Studien führt die Conjoint-Analyse zu signifikant besseren Ergebnissen als die ACA, obwohl 8, 9 bzw. 17 Eigenschaften berücksichtigt wurden. Pullman/Dodson/Moore (1999) können weiterhin zeigen, daß die Voll-Profil-Methode mit allen 17 Eigenschaften einer individualisierten Form, bei der lediglich die individuell wichtigsten 8 Eigenschaften bei der Stimulibeschreibung integriert werden, mit einer einzigen Ausnahme überlegen ist (vgl. Pullman/Dodson/Moore (1999), S. 131).

Tabelle 2-5: Empirische Vergleiche zwischen hybriden Verfahren und der Self-Explicated-Methode (Teil 1)

Studie	Verglichene Methoden	Gütemaße	Ergebnisse	Produkte	Anzahl Eigenschaften	Stichprobe	Experimentdesign
Traditionelle hybride Verfahren vs. Self-Explicated-Methode							
Green/Goldberg/ Montemayor (1981)	Hybrid Self-Explicated	Interne Validität	**Hybrid sig. besser als Self-Explicated**	Antibiotikum	8	404	within
Goldberg (1980) (zitiert nach Green/Goldberg/ Montemayer (1981))	Hybrid Self-Explicated	Prognosevalidität	**Hybrid besser als Self-Explicated a)**	Telekommunikationszubehör	7	324	within
Cattin/Hermet/ Pioche (1982)	Hybrid Self-Explicated	Prognosevalidität	keine Unterschiede a)	Bankservice	5	42	within
Green/Goldberg/ Wiley (1982)	Hybrid Self-Explicated	Prognosevalidität	**Hybrid sig. besser als Self-Explicated**	neues Haushaltsgerät	7	476	within
Akaah/Korgaonkar (1983)	Hybrid Self-Explicated	Prognosevalidität	keine sig. Unterschiede	HMO Pläne	6	80	within
Green/Krieger (1996)	Hybrid Self-Explicated	Prognosevalidität	**Hybrid sig. besser als Self-Explicated**	Telefon	15	600	within
Pullman/Dodson/ Moore (1999)	Hybrid Self-Explicated	Prognosevalidität	Ergebnisse variieren über Kriterien	Nationalpark	17	390 b)	between
Huber/Danesghar/ Ford (1971)	Hybrid Self-Explicated	Externe Validität	Self-Explicated besser als Hybrid a) c)	Arbeitsplatzwahl	5	30 b)	within
Adaptive hybride Verfahren vs. Self-Explicated-Methode							
Srinivasan/Park (1997)	CCA Self-Explicated	Externe Validität	keine sig. Unterschiede	Arbeitsplatzwahl	8	121 b)	within

Tabelle 2-5: Empirische Vergleiche zwischen hybriden Verfahren und der Self-Explicated-Methode (Teil 2)

Studie	Verglichene Methoden	Gütemaße	Ergebnisse	Produkte	Anzahl Eigenschaften	Stichprobe	Experimentdesign
Computergestützte adaptive Verfahren vs. Self-Explicated-Methode							
Green/Krieger/ Agarwal (1993)	ACA Self-Explicated	Reliabilität	keine sig. Unterschiede	Auto	8	133 b)	within
Agarwal/Green (1991)	ACA Self-Explicated	Prognosevalidität	Self-Explicated sig. besser als ACA d)	Appartment	6	170 b)	within
Green/Schaffer/ Patterson (1991)	ACA Self-Explicated	Prognosevalidität	**ACA besser als Self-Explicated a)**	Auto	8	96	within
Green/Krieger/ Agarwal (1993)	ACA Self-Explicated	Prognosevalidität	Ergebnisse variieren über Kriterien	Auto	8	133 b)	within
Huber et al. (1993)	ACA Self-Explicated	Prognosevalidität	**ACA sig. besser als Self-Explicated**	Kühlschrank	5 bzw. 9	393	within
Pullman/Dodson/ Moore (1999)	ACA Self-Explicated	Prognosevalidität	Ergebnisse variieren über Kriterien	Nationalpark	17	390 b)	between
Hensel-Börner/ Sattler (2000)	ACA Self-Explicated	Prognosevalidität	keine sig. Unterschiede	Kaffee	8	144	within und between
		Externe Validität	keine Unterschiede a)				
Hensel-Börner/ Sattler (2000)	CCC Self-Explicated	Prognosevalidität	keine sig. Unterschiede	Kaffee	8	144	within und between
		Externe Validität	Ergebnisse variieren über Kriterien				

a) keine Aussage zur Signifikanz; b) Befragt wurden Studierende
c) Für „inexperienced" Probanden gibt es keine Unterschiede zwischen den beiden Methoden.
d) Ausnahme: Die ACA ist dem reinen SE-Teil des Interiews signifikant überlegen (vgl. (Agarwal/Green (1991), S. 144).
Quelle: Eigene Erstellung in teilweiser Anlehnung an Sattler/Hensel-Börner (2000), S. 128 f.

Vergleicht man das Abschneiden der traditionellen hybriden und der adaptiven Varianten gegenüber der Self-Explicated-Methode separat, ergibt sich ein leicht differenziertes Bild. Während noch bei 50% (4 von 8) der Vergleiche die traditionellen hybriden Verfahren zu besseren Ergebnisse führen als die Self-Explicated-Methode, trifft dies bei den adaptiven Verfahren in lediglich 18% der Gegenüberstellungen zu. Die überwiegende Mehrzahl der Vergleiche zeigt, daß die Self-Explicated-Methode zu vergleichbaren, teilweise sogar besseren Ergebnissen führt als die verschiedenen adaptiven Conjoint-Verfahren. Damit schneiden die traditionellen hybriden Verfahren im Vergleich zur Self-Explicated-Methode relativ gesehen besser ab als die adaptiven Verfahren. Dies wirft die Frage auf, ob methodische Schwachstellen Gründe für die bisher häufige Unterlegenheit der ACA sind. Ob z.B. der Einsatz von Paarvergleichen die entsprechenden Resultate maßgeblich beeinflußt, kann durch Vergleiche zu den anderen adaptiven Verfahren nur bedingt analysiert werden. Zwar müssen in der CCA von Srinivasan/Park (1997) alle Stimuli simultan bewertet werden, allerdings handelt es sich hierbei nicht um eine computergestützte Befragung. Durch die 2-wöchige Zeitverzögerung zwischen Self-Explicated- und Conjoint-Teil sind systematische Verzerrungen nicht auszuschließen und ein direkter Vergleich nicht zulässig. Einen ersten empirischen Vergleich zwischen der CCC und der ACA liefern Hensel-Börner/Sattler (2000). Dabei schneidet die CCC (teilweise signifikant) besser ab als die ACA.[150] Diese bisher einzige empirische Überprüfung der beiden Varianten kann jedoch noch keine Antwort auf die Frage geben, ob mit alternativen Verfahren zur ACA validere Ergebnisse bei der Prognose von Kaufverhalten erzielt werden können. Vielmehr zeigen auch diese Befunde deutlich, daß bei der folgenden empirischen Überprüfung neuer adaptiver Verfahren die Self-Explicated-Methode einbezogen werden muß.

2.6 Zusammenfassung

Ziel der vorangestellten Ausführungen war, zunächst einen Überblick über verschiedene Ansätze zur Präferenzmessung zu geben. Dabei wurden alle bisher bekannten Verfahren kurz dargestellt und systematisiert. Grundsätzlich wurden 3 verschiedene Verfahrensbereiche zur Messung von Präferenzen differenziert. Die Self-Explicated-Verfahren als direkte Befragungsmethode und im Gegensatz dazu die dekompositionellen (indirekten) Verfahren. Bedeutendster Vertreter der

---

[150] Vgl. Hensel-Börner/Sattler (2000), S. 13 ff.

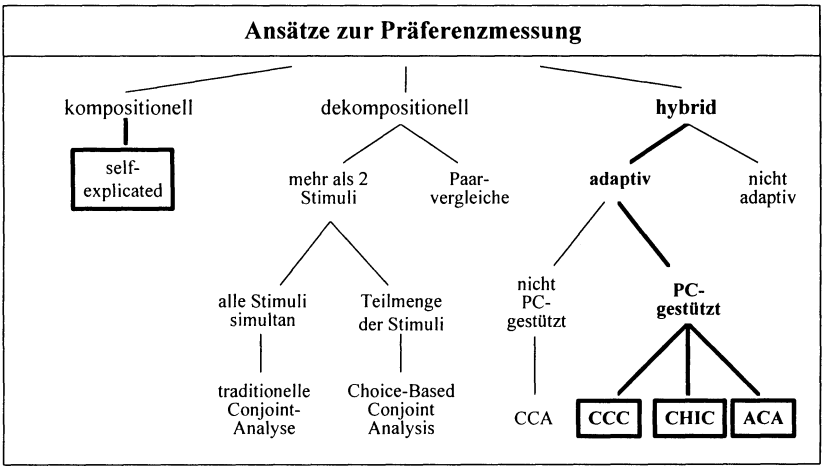
indirekten Verfahren ist die Conjoint-Analyse. Verschiedene Varianten der Conjoint-Analyse wurden hinsichtlich der Anzahl simultan zu beurteilender Stimuli und der Form der Präferenzabfrage systematisiert. Neben diesen beiden gegensätzlichen Vorgehensweisen zur Messung von Nachfragerpräferenzen hat sich mit den hybriden Ansätzen, einer Mischform aus beiden Verfahren, ein dritter großer Verfahrenszweig etabliert. Besondere Bedeutung kommt hierbei der ACA als bisher einzig computergestützte adaptive Methode zu.

Neben der Diskussion der jeweiligen Vor- und Nachteile alternativer Verfahren sind die Befunde bisheriger empirischer Vergleiche zusammengefaßt worden. Vergleicht man bisherige Forschungsergebnisse zur Reliabilität, vor allem aber im Hinblick auf die Validität verschiedener Varianten der rein dekompositionellen Conjoint-Analysen, wird deutlich, daß keiner der bestehenden Formen eine grundsätzliche Überlegenheit zugesprochen werden kann. Studien, die Vergleiche zwischen den 3 Verfahrenszweigen durchführen, können zudem die in der praktischen Anwendung dominante Rolle der ACA nicht stützen. Daher stellt sich die Frage, ob dies mit methodischen Schwächen des Verfahrens zu begründen ist. Die ACA als bisher einzig computergestützte hybride Conjoint-Analyse setzt im dekompositionellen Verfahrensteil Paarvergleiche ein. Damit wird lediglich eine Variante dekompositionellen Vorgehens im Bereich computergestützter hybrider Conjoint-Analysen berücksichtigt, die zudem bei rein dekompositionellen Conjoint-Analysen eine eher untergeordnete Rolle spielt.[151] Ziel der vorliegenden Arbeit ist daher, den Bereich computergestützter Verfahren um neue Methoden zu erweitern, die im Conjoint-Teil entweder alle oder eine Teilmenge der zu bewertenden Stimuli simultan präsentieren. Damit wird der Vielfalt der Gestaltungsformen dekompositioneller Varianten bei computergestützten Conjoint-Analysen Rechnung getragen, wobei auch neuere Entwicklungen wie die Choice-Based Conjoint Analysis integriert werden. Abbildung 2-6 gibt noch einmal einen zusammenfassenden Überblick über alle diskutierten Ansätze zur Präferenzmessung und verdeutlicht die Einordnung der neu entwickelten Verfahren.

---

[151] Vgl. dazu Wittink/Cattin (1989) S. 93 als auch Wittink/Vriens/Burhenne (1994), S. 44. Des weiteren wird dies im Rahmen des Literaturüberblicks in Kapitel 2.2.2.2 deutlich. Tabelle 2-1 zeigt, daß nur bei 4 der 17 ausgewerteten Studien Paarvergleiche zum Untersuchungsgegenstand zählten.

Abbildung 2-6: Überblick über alle diskutierten Ansätze zur Präferenzmessung



Quelle: Eigene Erstellung in teilweiser Anlehnung an Green/Srinivasan (1990), S. 9.

Ein weiteres wichtiges Ergebnis der empirischen Vergleiche betreffen die Befunde für die Self-Explicated-Methode. Auch wenn in einem theoretischen Vergleich mit der Conjoint-Analyse zahlreiche Argumente für den Einsatz dieser Methode sprechen, zeigt sich diese Überlegenheit in empirischen Studien bisher nicht. Weder die rein dekompositionelle Conjoint-Analyse noch hybride Verfahren (sowohl traditionelle als auch adaptive) können allein auf Basis ihrer theoretischen Vorteile der einfacheren Self-Explicated-Methode generell als überlegen angesehen werden. Vielmehr schneidet das Self-Explicated-Verfahren vor allem im Hinblick auf Prognosevalidität und externer Validität in den vorliegenden Untersuchungen empirisch überraschend gut ab. Somit werden bei dem im Rahmen dieser Arbeit vorgenommenen empirischen Vergleich neuer adaptiver Methoden auch die Self-Explicated-Methode berücksichtigt (vgl. Hervorhebung in Abbildung 2-6), die unter Anwendungsgesichtspunkten wie Zeit- und Kostenaufwand den umfangreicheren Interviews aller Conjoint-Varianten vorzuziehen ist.

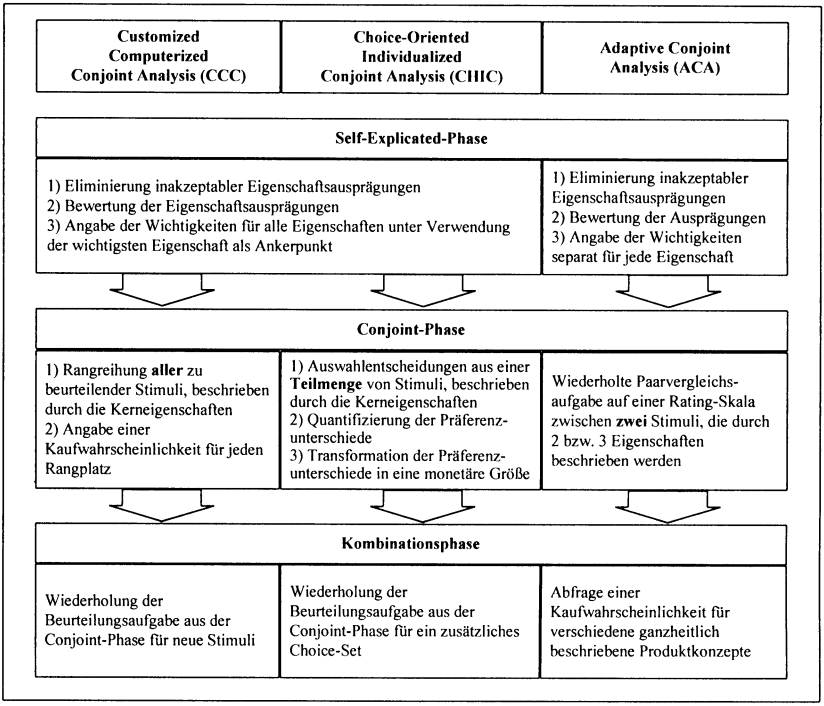
3 Untersuchte computergestützte hybride Verfahren

Nachdem in Kapitel 2 ein Überblick über Ansätze zur Präferenzmessung gegeben wurde, befaßt sich das folgende Kapitel mit einer ausführlichen Darstellung der computergestützten hybriden Verfahren CCC, CHIC und ACA, die im Mittelpunkt der vorliegenden Arbeit stehen. Die Einordnung dieser Ansätze wurde bereits im vorangestellten Methodenüberblick vorgenommen. Bei allen 3 untersuchten Methoden handelt es sich um hybride Verfahren, die einen kompositionellen mit einem dekompositionellen Befragungsteil verbinden. Des weiteren zeichnen sich die Verfahren durch ihren adaptiven Charakter aus, wodurch insbesondere der dekompositionelle Teil an vorherige Antworten individuell angepaßt wird. Voneinander abgegrenzt werden können die Ansätze durch die unterschiedliche Gestaltung dieses individualisierten Conjoint-Teils und das teilweise davon beeinflußte Ausmaß der Anpassung. Entsprechend der Strukturierung der rein dekompositionellen Ansätze nach der Anzahl simultan zu beurteilender Stimuli unterscheiden sich auch die CCC, CHIC und die ACA hinsichtlich dieses Merkmals. Bei der CCC werden im dekompositionellen Teil des Interviews alle zu bewertenden Stimuli gleichzeitig vorgelegt. In Anlehnung an die Choice-Based Conjoint Analysis werden den Probanden bei der CHIC jeweils verschiedene Sets (Teilmengen) von Stimuli präsentiert, aus denen u.a. eine Auswahlentscheidung abgefragt wird. Die ACA basiert im dekompositionellen Interviewteil auf Paarvergleichen, die mit unterschiedlichen Stimuli mehrfach wiederholt werden. Damit kann jedes dieser drei Verfahren einer anderen Kategorie der dekompositionellen Ansätze zugeordnet werden. Neben dieser im wesentlichen konzeptionellen Unterscheidung variieren die Verfahren auch in der Gestaltung der einzelnen Befragungsschritte. Die jeweiligen Besonderheiten der Ansätze werden in den folgenden Ausführungen erläutert.

Während die ACA ein sehr weit verbreiteter Ansatz ist, der sich in der Praxis mittlerweile als ein Standardinstrument zur Präferenzmessung etabliert hat, handelt es sich bei der CCC und insbesondere der CHIC um im Rahmen dieser Arbeit neu entwickelte Verfahren. Alle drei Verfahren werden im folgenden ausführlich dargestellt. Die Ausführungen werden dabei entsprechend der einzelnen Interviewschritte untergliedert. So werden jeweils zunächst die kompositionelle

als auch dekompositionelle Phase beschrieben. Die sich anschließende dritte Phase ist Grundlage für die Kombination der gewonnenen Informationen. Sowohl die Form der Datenerhebung als auch die Schätzung der Teilnutzenfunktionen werden in jeder Phase dokumentiert.

Abbildung 3-1: Interviewablauf der untersuchten computergestützten hybriden Verfahren im Vergleich



Quelle: Eigene Erstellung

Die Reihenfolge der Darstellung ist an deren Einordnung angesichts der dekompositionellen Phase angelehnt. Danach wird zunächst die CCC (alle Stimuli simultan), anschließend die CHIC (Teilmenge an Stimuli) und als drittes die ACA (zwei Stimuli simultan) dargestellt.

3.1 Customized Computerized Conjoint Analysis (CCC)

Grundlage der CCC bildet die von Srinivasan und Park (1997) entwickelte Customized Conjoint Analysis (CCA), bei der sich an die Self-Explicated-

Befragung eine Ranking- oder Rating-Aufgabe bezüglich aller zu bewertender Produkte anschließt. Das Design der Produktkonzepte wird dabei an die Antworten im Rahmen der direkten Befragung angepaßt und für jeden Probanden individuell gestaltet. Da die CCC in ihrer Konzeption auf der CCA aufbaut, entspricht die folgende Darstellung im wesentlichen den Ausführungen von Srinivasan/Park (1997). Die bei der CCC vorgenommenen Modifikationen werden jeweils an entsprechender Stelle erläutert.

3.1.1 Self-Explicated-Phase

3.1.1.1 Datenerhebung

Die Self-Explicated-Phase der CCC entspricht der von Srinivasan (1988) vorgestellten Conjunctive-Compensatory Self-Explicated-Methode.[152] Die Datenerhebung besteht dabei aus 3 aufeinander folgenden Befragungsschritten.

1. Schritt: Eliminierung inakzeptabler Eigenschaftsausprägungen

Der Proband hat die Möglichkeit, für ihn völlig inakzeptable Ausprägungen bei den verschiedenen Eigenschaften zu eliminieren.[153] Diese Ausprägungen werden in folgenden Schritten der Befragung nicht mehr berücksichtigt.[154]

2. Schritt: Erfassung von Präferenzen zu den Eigenschaftsausprägungen

Von den verbleibenden Ausprägungen muß der Proband für jede Eigenschaft die von ihm am meisten und am wenigsten präferierte Ausprägung identifizieren. Diese definieren die Extrempole einer 11-Punkte-Ratingskala (Präferenzwert 10 für die am meisten präferierte Ausprägung und 0 für die am wenigsten präferierte Ausprägung). Auf dieser Rating-Skala mit den für den Probanden besten und schlechtesten Ausprägung als Ankerpunkte, werden alle weiteren Ausprägungen je Eigenschaft eingeordnet.[155]

Die Bewertung der Eigenschaftsausprägungen (als auch der Eigenschaften im nächsten Schritt) erfolgt immer unter der Annahme, daß alle anderen Eigen-

---

[152] Vgl. hierzu Kapitel 2.1 dieser Arbeit und Srinivasan (1988), hier: S. 267 f.

[153] Vgl. Srinivasan/Park (1997), S. 287.

[154] In der derzeitigen Programmversion der CCC ist dieser Schritt noch nicht implementiert. Eine optionale Integration dieses Befragungsschritts in einer folgenden Programmversion ist geplant.

[155] Vgl. Srinivasan/Park (1997), S. 287.

schaften jeweils konstant sind. Folglich können vor allem bei quantitativen Eigenschaften die beste und schlechteste Ausprägung jeweils a priori festgelegt werden. So müßte z.B. bei Konstanz aller anderen Eigenschaftsausprägungen der niedrigste Preis am meisten und der höchste Preis am wenigsten von allen Probanden präferiert werden.[156] Die CCC bietet daher die Möglichkeit, die Abfrage nach bester und schlechtester Ausprägung für einzelne Eigenschaften auszuschließen. Die Ankerpunkte der Rating-Skala werden dann bereits vom Marktforscher fixiert und der Proband muß lediglich die „mittleren" Ausprägungen einordnen. Für dieses Vorgehen sprechen vor allem Überlegungen, die Effizienz der Datenerhebung zu erhöhen. Dabei ist Ziel, den Probanden nur mit solchen Fragen zu konfrontieren, bei denen, plausibles und konsistentes Verhalten vorausgesetzt, keine eindeutige Antwort erwartet werden kann. Auf eine Einordnung der dazwischen liegenden Ausprägungen wird bewußt nicht verzichtet, um mögliche Schwellenwerte oder Sprünge in den Teilnutzenfunktionen erfassen zu können.

3. Schritt: Abfrage der Wichtigkeiten der Eigenschaften

Im dritten Schritt der Self-Explicated-Phase werden die Wichtigkeiten der Eigenschaften erfragt. Ausschlaggebend dafür ist die Bedeutung der Verbesserung von der am wenigsten zu der am meisten präferierten Ausprägung bei jeder Eigenschaft. Damit wird die Wichtigkeit einer Eigenschaft explizit an die Ausprägungen in der sie auftritt gekoppelt und das Konzept der Wichtigkeit mit einer konkreteren Definition verbunden.[157] Gleichzeitig wird dabei die Gefahr des Bandbreiteneffekts verringert.[158] Um die Wahrscheinlichkeit zu erhöhen, die Wichtigkeiten der Eigenschaften tatsächlich auf einer Ratio-Skala zu erfassen, wird zunächst die für den Probanden insgesamt wichtigste Eigenschaft definiert.[159] Hierfür wird erfragt, bei welcher Eigenschaft die Verbesserung von der am wenigsten zu der meistpräferierten Ausprägung am wertvollsten ist. Dieser Eigenschaft wird

---

156 Ausnahmen können hierbei auftreten, sofern der Preis als Qualitätsindikator dient und der Informationspreiseffekt größer als der Allokationspreiseffekt ist. Ein solches Verhalten der Probanden würde sich in einer höheren Präferenz für einen höheren Preis widerspiegeln. Zur Separierung der Allokations- und Informationsfunktion des Preises vgl. Sattler/Rao (1997), hier: S. 1287.

157 Häufig wird die Wichtigkeit einer Eigenschaft in die Frage: „Wie wichtig ist Ihnen die Eigenschaft X?" erhoben. Bei einem solchen Vorgehen ist oft nicht eindeutig, wie das Konzept der Wichtigkeit zu interpretieren ist und birgt die Gefahr, daß Probanden diese Frage unterschiedlich interpretieren (vgl. Srinivasan (1988), S. 296).

158 Zur Problematik des Bandbreiteneffekts vgl. Sattler/Gedenk/Hensel-Börner (2000).

159 Vgl. Srinivasan (1988), S. 296.

der Wert 10 zugeordnet und dient für die Beurteilung der anderen Eigenschaften als Ankerpunkt („critical attribute"). Die Wichtigkeit aller weiteren Eigenschaften ist daraufhin durch die Auskunftspersonen auf einer Rating-Skala 0 bis 10 zu beurteilen. Diese Skala ist einseitig mit der für den Probanden als am wichtigsten beurteilten Eigenschaft bei dem Wert 10 verankert. Je höher der Wert, den der Proband einer Eigenschaft zuordnet, desto wichtiger ist sie für ihn.[160]

3.1.1.2 Schätzung der Teilnutzenwerte

Entsprechend dem multiattributiven Präferenzmodell werden auf Basis der Befragungsschritte 1 bis 3 die Self-Explicated-Teilnutzenwerte für alle J Eigenschaften berechnet. Zur Vereinfachung erfolgt zunächst eine Umskalierung der Wichtigkeiten in der Form, daß deren Summe 100 ergibt. Der Vorteil dieser Umskalierung liegt darin, daß die Spannbreite der ermittelten Teilnutzenwerte je Eigenschaft direkt als relatives Bedeutungsgewicht interpretiert werden kann. Die Self-Explicated-Teilnutzenwerte für jede Eigenschaftsausprägung erhält man durch Multiplikation der umskalierten Wichtigkeiten der Eigenschaften mit dem normierten Präferenzwert der Ausprägung:

$$\text{(3-1)}\quad P_{ijk}^{SE} = G_{ij}\left(D_{ijk} / 10\right) \qquad (i \in I;\ j \in J;\ k \in K_j)$$

wobei:

P_{ijk}^{SE}: Self-Explicated (SE) Teilnutzen der k-ten Ausprägung bei der j-ten Eigenschaft für den i-ten Probanden,

G_{ij}: Umskalierte Wichtigkeit der j-ten Eigenschaft beim i-ten Probanden,

D_{ijk}: Präferenzwert der k-ten Ausprägung bei der j-ten Eigenschaft beim i-ten Probanden,

K_j: Indexmenge der Eigenschaftsausprägungen der j-ten Eigenschaft,

J: Indexmenge der Eigenschaften,

I: Indexmenge der Probanden.

---

[160] Vgl. Srinivasan/Park (1997), S. 287.

Nach Berechnung der Teilnutzenwerte werden diese so normiert, daß ihre Summe über alle Ausprägungen je Eigenschaft 0 beträgt.[161]

Von allen bewerteten J Eigenschaften werden die aus Sicht des i-ten Probanden C_i ($C_i \in J$) wichtigsten Eigenschaften ausgewählt. Dies sind die sog. Kerneigenschaften, die in der anschließenden Conjoint-Aufgabe berücksichtigt werden. Durch die individuell verschiedenen Wichtigkeiten der Eigenschaften variieren die im Conjoint-Interview verwendeten Kerneigenschaften von Proband zu Proband. Diese individuelle Anpassung der Kerneigenschaft an die Antworten der Self-Explicated-Phase bestimmt den adaptiven Charakter der Methode und hat zur Bezeichnung **Customized Conjoint Analysis** geführt.[162] Da alle zu bewertenden Stimuli im dekompositionellen Befragungsteil simultan präsentiert werden, ist eine Anpassung der Stimulikonstruktion an vorangegangene Antworten der Auskunftsperson lediglich einmalig möglich. Aus dieser Konzeption des Conjoint-Interviews ergibt sich folglich ein geringeres Maß an Adaption als beispielsweise bei der ACA.

Die Auswahl der Kerneigenschaften und die individuelle Anpassung der Conjoint-Phase erfolgt bei der CCA in einem zweiten Interview. Da es sich bei der CCA um eine schriftliche Befragung handelt, kann zwischen diesen beiden Interviewphasen eine erhebliche Zeitverzögerung entstehen (in einer empirischen Studie von Srinivasan/Park (1997) waren es zwei Wochen).[163] Als Folge davon ist u.a. mit erhöhtem Erhebungsaufwand und einem Ausfall von Versuchspersonen zu rechnen. Gleichzeitig besteht die Möglichkeit, daß sich die Präferenzen der Probanden über die Zeit verändert haben und die Validität der Ergebnisse dadurch eingeschränkt wird. Um diesen wesentlichen Nachteil der CCA zu überwinden, basiert die **Customized Computerized Conjoint Analysis (CCC)** auf einem vollständig computergestützten Interview, bei dem eine unmittelbare Anpassung des Conjoint-Teils an die Antworten des Self-Explicated-Teils erfolgt. Alle notwendigen Daten werden dabei innerhalb eines Interviews mit dem Probanden erhoben.

---

161 Vgl. Srinivasan/Park (1997), S. 287.
162 Vgl. Srinivasan/Park (1997), S. 286.
163 Vgl. Srinivasan/Park (1997), S. 288.

3.1.2 Conjoint-Phase

In Anlehnung an die Systematisierung dekompositioneller Ansätze in Kapitel 2 wird die Conjoint-Phase der computergestützten hybriden Verfahren untergliedert. Ein erstes Kennzeichen ist wie bereits erwähnt die Anzahl simultan zu beurteilender Stimuli. Teilweise in Abhängigkeit davon, ergeben sich die Gestaltungsmöglichkeiten für die Abfrage der Präferenz bzw. Form der Beurteilungsaufgabe. Als dritter Unterpunkt wird jeweils angeführt, wie auf Basis der erhobenen Präferenzdaten die Conjoint-Teilnutzenwerte geschätzt werden.

3.1.2.1 Anzahl simultan zu bewertender Stimuli

Die Anzahl der Stimuli, die den Probanden zur Bewertung vorgelegt werden, wird durch die beschreibenden Eigenschaften und deren Ausprägungen determiniert. Zur Beschreibung der Stimuli werden die in der Self-Explicated-Phase identifizierten Kerneigenschaften herangezogen. Um die Probanden kognitiv nicht zu überfordern, sollten nicht mehr als 6 Eigenschaften als Kerneigenschaften zugelassen werden.[164] Zudem wird die Zahl der Ausprägungen je Kerneigenschaft auf (maximal) 3 begrenzt. Der Vorteil dieser Reduzierung besteht darin, daß damit für alle Eigenschaften nahezu gleich viele Ausprägungen betrachtet werden und Verzerrungen aufgrund des sog. Number-of-Level-Effekts[165] vermieden werden können.[166] Falls eine Kerneigenschaft mehr als 3 Ausprägungen aufweist, wird jeweils eine Auswahl nach dem gleichen Grundprinzip vorgenommen. Mit der am wenigsten und am meisten präferierten Ausprägung wird zunächst die Bandbreite einer Eigenschaft definiert. Als dritte Ausprägung wird eine mittlere, die einen Wert nahe 5 auf der 11-Punkte-Rating-Skala erhalten hat, ausgewählt.[167]

---

[164] Vgl. hierzu auch die Diskussion zur Anzahl an Eigenschaften bei dekompositionellen Ansätzen zur Präferenzmessung in Kapitel 2.3.1.

[165] Der Number-of-Level-Effekt besagt, daß je mehr Ausprägungen bei einer Eigenschaft berücksichtigt werden, desto höher wird deren Bedeutungsgewicht. Empirische Belege dafür liefern u.a. Wittink/Krishnamurthi/Nutter (1982) und Wittink/Krishnamurthi/Reibstein (1989). Steenkamp/Wittink (1994) zeigen des weiteren, daß der Number-of-Level-Effekt auch unabhängig vom Skalenniveau der Beurteilungsaufgabe auftritt.

[166] Vgl. Srinivasan/Park (1997), S. 287.

[167] Falls eine Eigenschaft mit lediglich 2 Ausprägungen auftritt, werden diese beiden im Rahmen der Stimulikonstruktion berücksichtigt.

Obwohl dadurch die Anzahl an Eigenschaften und deren Ausprägungen für die Conjoint-Phase des Interviews begrenzt werden, ist eine Bewertung aller möglichen Eigenschaftskombinationen nicht möglich. Bei nur 4 Kerneigenschaften mit jeweils 3 Ausprägungen müßten den Probanden bereits 81 Produktkonzepte vorgelegt werden. Das simultan zu bewertende Stimuli-Set wird daher mittels eines fraktionierten faktoriellen Designs erstellt, wodurch das vollständige Design möglichst gut repräsentiert wird.[168] Zielkriterium in diesem Fall ist das Prinzip der Orthogonalität[169], bei dem die Eigenschaften und deren Ausprägungen unkorreliert sind.[170] Da sowohl die Kerneigenschaften als auch deren Ausprägungen über die Probanden variieren, unterscheiden sich die Designpläne zwischen den Probanden. Die damit verbundene Individualisierung der Stimuli bestimmt maßgeblich den adaptiven Charakter der CCC.[171] Durch die individualisierten Stimuli werden gleichzeitig systematische Einflüsse auf die Ergebnisse durch die Stimulibeschreibung, wie z.B. die Reihenfolge, in der die Eigenschaften aufgeführt sind, vermieden.[172]

Als letzte Einflußgröße auf die Anzahl zu bewertender Stimuli ist die Menge zu schätzender Parameter zu berücksichtigen. Je nach Anzahl an Eigenschaften und angenommener Funktionsverläufe für die Teilnutzenfunktionen sind unterschiedlich viele Parameter zu schätzen. Die Wahl der Funktionsverläufe wird maßgeblich von der Art der Eigenschaften determiniert. Für qualitative bzw. kategoriale Eigenschaften wie z.B. „Marke“ wird üblicherweise ein Teilnutzenmodell angenommen, bei dem für jede Ausprägung ein separater Teilnutzenwert zu schätzen ist.[173] Im Gegensatz dazu kann bei quantitativen Eigenschaften ein Vektormodell und damit ein linearer Zusammenhang zwischen den Eigenschaftsausprägungen unterstellt werden. In diesem Fall ist die Schätzung eines Parameterwertes ausrei-

---

168 Zu den Problemen bei einem vollständigen Design vgl. auch Kapitel 2.2.1.1.

169 Zum Begriff der Orthogonalität und seiner formalen Definition vgl. Stallmeier (1993), S. 144.

170 Zur Entwicklung von fraktionierten faktoriellen Designs vgl. Addelman (1962).

171 Vgl. Srinivasan/Park (1997), S. 287. Die Autoren sprechen in diesem Zusammenhang nicht von einem adaptiven, sondern einem „customized“ Design. Zur Begriffsabgrenzung im Rahmen dieser Arbeit vgl. Kapitel 2.4.

172 Dieser Einflußfaktor auf die Ergebnisse einer Conjoint-Analyse wird in der Literatur als „Positionseffekt“ bzw. „order effects“ bezeichnet. Empirische Belege für deren Einfluß liefern z.B. Acito (1977), S. 205; Chrzan (1994), S. 169 f. und Perry (1996), S. 111.

173 Vgl. Green/Srinivasan (1978), S. 106.

chend.[174] Insgesamt muß bei der Entscheidung über die Anzahl zu erhebender Datenpunkte bei der CCC aber gleichzeitig sichergestellt werden, daß alle Stimuli simultan auf dem Bildschirm darstellbar sind.

3.1.2.2 Form der Beurteilungsaufgabe

Ein wesentliches Kennzeichen der CCC liegt darin, daß die Probanden alle Stimuli der Conjoint-Phase simultan zu bewerten haben. Die Präferenzen können dabei entweder in Form einer Rangreihung oder mit Hilfe einer Rating-Skala erhoben werden.[175] Bei der CCC werden diese beiden Beurteilungsaufgaben miteinander kombiniert. Die Auskunftsperson wird zunächst gebeten, die Stimuli in eine Präferenzrangfolge zu bringen. Zusätzlich wird im Anschluß daran für jeden Rangplatz die Kaufwahrscheinlichkeit auf einer Skala von 0 bis 100% erfragt. Zur Erleichterung dieser Aufgabe und um inkonsistente Antworten zu vermeiden, werden die Produkte vor dieser Abfrage gemäß der von den Probanden angegebenen Präferenzrangfolge sortiert. Mit der Kombination dieser beiden Beurteilungsaufgaben werden die Vorteile der jeweiligen Vorgehensweise miteinander verknüpft. Während die Angabe einer Präferenzrangfolge einen echten Trade-Off zwischen verschiedenen Eigenschaften fordert, werden über die Kaufwahrscheinlichkeiten die Präferenzunterschiede zwischen den Stimuli quantifiziert. Gleichzeitig kann damit erfaßt werden, ob überhaupt eine Kaufwahrscheinlichkeit für die jeweiligen Produktkonzepte vorliegt. Gleichzeitig werden somit über die Präferenzdaten hinaus auch Informationen hinsichtlich des Entscheidungsverhaltens der Probanden erhoben.[176] Eine solche Kombination der beiden Beurteilungsaufgaben schlagen beispielsweise auch Green/Krieger (1996) vor.[177] Zweifelsohne wird dadurch der Erhebungsaufwand insgesamt erhöht. Ein empirischer Vergleich der Ergebnisse jeweils auf Basis der einzelnen Bewertungsschritte separat, kann zeigen, ob dieser zusätzliche Erhebungsaufwand durch validere Ergebnisse gerechtfertigt ist.

---

174 Zu den unterschiedlichen Teilnutzenmodellen vgl. Green/Srinivasan (1978), S. 105 f.

175 Srinivasan/Park (1997) sprechen bei der Beschreibung der CCA davon, daß grundsätzlich Ranking- oder Rating-Daten erhoben werden können (S. 287). In der empirischen Untersuchung wurden die Präferenzen durch Rangreihung der Stimuli erfragt.

176 Vgl. hierzu auch Hahn (1997), der als wesentlichen Kritikpunkt an der traditionellen Conjoint-Analyse mit Rangreihung anführt, daß lediglich Präferenzen und keine Kaufentscheidungsdaten erhoben werden (Hahn (1997), hier: S. 41).

177 Vgl. Green/Krieger (1996), S. 853.

3.1.2.3 Schätzung der Teilnutzenwerte

Auf Basis der erhobenen Präferenzen werden für alle Kerneigenschaften Teilnutzenwerte bestimmt. Je nach Skalenniveau der abhängigen Variable (also der Form der Beurteilungsaufgabe) werden die Teilnutzenwerte der Conjoint-Phase mit Hilfe unterschiedlicher Schätzverfahren ermittelt. Sofern die Stimuli auf metrischem Skalenniveau (Rating-Skala) beurteilt werden, wird eine OLS-Schätzung (Ordinary Least Squares) zur Ermittlung der Teilnutzenwerte durchgeführt. Im Falle einer Präferenzabfrage in Form von Rangreihung, muß streng genommen auch ein nicht-metrisches Verfahren eingesetzt werden. In der vorliegenden Untersuchung wurde das Verfahren LINMAP (Linear Programming Techniques for Multidimensional Analysis of Preferences)[178] angewendet, das auf Basis einer linearen Optimierung die Teilnutzenwerte bei ordinal skalierten Präferenzwerten ermittelt.[179] Um die ermittelten Conjoint-Teilnutzen der Kerneigenschaften mit denen aus der Self-Explicated-Phase vergleichen zu können, werden diese in zweierlei Hinsicht umskaliert. Zum einen werden die Teilnutzenwerte so normiert, daß die Summe innerhalb jeder Kerneigenschaft Null ergibt und zum anderen so umskaliert, daß die Summe der Conjoint-Bedeutungsgewichte der Kerneigenschaften gleich der Summe der Self-Explicated-Bedeutungsgewichte der Kerneigenschaften ist.

3.1.3 Kombinationsphase

Neben den Self-Explicated-Teilnutzenwerten für alle untersuchten Eigenschaften liegen zusätzlich umskalierte Conjoint-Teilnutzenwerte für die jeweiligen Kerneigenschaften eines Probanden vor. Die Informationen bezüglich der Kerneigenschaften werden folglich mit zwei unterschiedlichen Methoden erhoben. Die Grundidee der CCC ist, die Informationen über die Kerneigenschaften aus beiden

---

178 Zur ausführlichen Erläuterung des Algorithmus vgl. Srinivasan/Shocker (1973).

179 Obwohl Ranking-Daten nicht das notwendige Skalenniveau aufweisen, werden die Teilnutzenwerte häufig auch mit Hilfe einer OLS-Schätzung ermittelt, wobei die Präferenzdaten als metrisch interpretiert werden. Vorangegangene Studien haben gezeigt, daß zwischen metrischen und nicht-metrischen Präferenzdaten bei der Anwendung von Conjoint-Analysen in Form von OLS-Schätzungen keine wesentlichen Unterschiede auftreten bzw. robuste Schätzergebnisse erzielt werden (vgl. Green/Srinivasan (1978); Carmone/Green/Jain (1978); Jain/Acito/Malhotra/Mahajan (1979) Wittink/Cattin (1981)). Der Einsatz einer OLS-Schätzung zur Auswertung von Präferenzdaten, die in Form von Rangreihung erhoben wurden, kann daher als zulässig angesehen werden. Dieses Vorgehen hat den Vorteil, daß OLS-Schätzungen mittels üblichen Softwareprogrammen wie z.B. SPSS oder Excel durchgeführt werden können.

Phasen optimal zu kombinieren, um die tatsächlichen Präferenzstrukturen der Probanden bestmöglich abzubilden. Grundlage hierfür ist ein sog. Validation-Set. Dieses beschreibt die verschiedenen Produktkonzepte ebenfalls durch die individuellen Kerneigenschaften, unterscheidet sich aber in der Zusammensetzung der verschiedenen Eigenschaftsausprägungen von denen der Conjoint-Phase, so daß die Produktprofile so weit wie möglich von den bereits beurteilten abweichen. Auch für das Validation-Set, das aus weniger Stimuli bestehen kann als die eigentliche Conjoint-Aufgabe, werden von jedem Probanden Präferenzen erfragt.[180] Die Form der Präferenzabfrage entspricht dabei dem Vorgehen in der Conjoint-Phase.

Die Self-Explicated- und Conjoint-Teilnutzenwerte werden in der Form kombiniert, daß die auf Basis der kombinierten Teilnutzenwerte prognostizierten Präferenzen für das Validation-Set möglichst gut mit den tatsächlich beobachteten übereinstimmen. Um diese optimale Kombination zu bestimmen, wird zunächst eine Gewichtung der beiden Teilnutzenwerte für die Kernattribute vorgenommen. Für die gewichteten Teilnutzenwerte gilt:

$$(3\text{-}2) \qquad P_{ijk}^{w} = w_i P_{ijk}^{CA} + (1 - w_i) P_{ijk}^{SE} \qquad (i \in I;\ j \in C_i;\ k \in K_j)$$

wobei:

P_{ijk}^{w} Gewichteter Teilnutzen der k-ten Ausprägung bei der j-ten Kerneigenschaft für den i-ten Probanden,

w_i: Gewichtungsfaktor beim i-ten Probanden, wobei $(0<w_i<1)$,

P_{ijk}^{CA}: Umskalierter Conjoint-Teilnutzen der k-ten Ausprägung bei der j-ten Kerneigenschaft für den i-ten Probanden,

P_{ijk}^{SE}: Self-Explicated-Teilnutzen der k-ten Ausprägung bei der j-ten Kerneigenschaft für den i-ten Probanden,

C_i: Indexmenge der Kernattribute beim i-ten Probanden.

Für alternative Gewichte w_i (w_i = 0; 0,01; 0,02; ...;1) wird jeweils die Übereinstimmung zwischen prognostizierten und tatsächlichen Präferenzen für das Vali-

---

[180] Srinivasan/Park (1997) schlagen beispielsweise 6 Stimuli für das Validation-Set vor (hier: S. 288).

dation-Set berechnet. Die optimale Gewichtung ist für den Wert w_i erreicht, bei dem sich die höchste Übereinstimmung ergibt. Gemessen wird diese je nach Skalenniveau der erhobenen Präferenzwerte durch den Korrelationskoeffizient nach Pearson oder Kendall.[181]

Wenn die optimale Gewichtung w_i^* für einen Probanden ermittelt wurde, kann seine Präferenz für einen neuen Stimulus prognostiziert werden. Hierfür werden unter Annahme eines additiven Nutzenmodells die optimal gewichteten Teilnutzenwerte (aus Conjoint- und Self-Explicated-Phase) für die Ausprägungen der Kerneigenschaften und die Self-Explicated-Teilnutzenwerte für die Ausprägungen der Nicht-Kerneigenschaften addiert. U_i ist dann die prognostizierte Präferenz, gemessen durch den prognostizierten Gesamtnutzenwert, des i-ten Probanden für ein Produkt[182]:

$$(3\text{-}3) \qquad U_i = \sum_{j \in CI} \left[w_i^* P_{ijk_j}^{CA} + \left(1 - w_i^*\right) P_{ijk_j}^{SE} \right] + \sum_{j \in NC_i} P_{ijk_j}^{SE} \qquad (i \in I;\ k \in K_j)$$

wobei:

U_i: Gesamtnutzenwert eines Produkts für den i-ten Probanden,

w_i^*: Optimaler Gewichtungsfaktor zwischen Conjoint-Teilnutzen und Self-Explicated-Teilnutzen bei den Kernattributen für den i-ten Probanden,

NC_i: Indexmenge der Nicht-Kernattribute beim i-ten Probanden.

Hinsichtlich dieser Berechnung der Gesamtnutzenwerte nimmt die CCC eine weitere Modifikation gegenüber der CCA vor. Die Motivation für die vorgenommene Erweiterung besteht in folgender Überlegung: In Gleichung (3-3) gehen die Nicht-Kerneigenschaften mit ihrem vollen Self-Explicated-Gewicht (Gewichtungsfaktor 1) in die prognostizierte Präferenz eines neuen Produkts ein, wohingegen die Kerneigenschaften mit der optimalen Gewichtung von Conjoint- und Self-Explicated-Gewichten berücksichtigt werden. Dadurch kann der Fall

---

181 Vgl. Srinivasan/Park (1997), S. 288.

182 Aus Gründen der Übersichtlichkeit wird in Gleichung (3-3) kein zusätzlicher Index für das Produkt integriert.

auftreten, daß ein oder mehrere Bedeutungsgewichte von Kerneigenschaften nach der optimalen Gewichtung (z.T. wesentlich) kleiner sind als die Bedeutungsgewichte von Nicht-Kerneigenschaften. Dies ergibt sich insbesondere dann, wenn einzelne Kerneigenschaften im Conjoint-Teil der Befragung als relativ unwichtig eingestuft werden und sich für die kombinierten Teilnutzenwerte ein relativ hohes Gewicht für die Conjoint-Teilnutzen insgesamt ergibt. Um diesem Effekt entgegenzuwirken, wird Gleichung (3-3) so modifiziert, daß auch die Nicht-Kerneigenschaften mit dem optimalen Gewicht ($1-w_i^*$) multipliziert werden:

$$(3\text{-}4) \qquad U_i = \sum_{j \in Ci} \left[w_i^* P_{ijk_j}^{CA} + \left(1 - w_i^*\right) P_{ijk_j}^{SE} \right] + \sum_{j \in NC_i} \left(1 - w_i^*\right) P_{ijk_j}^{SE} \qquad (i \in I;\ k \in K_j)$$

Mit dieser veränderten Berechnung der Gesamtnutzenwerte liegt die Wahrscheinlichkeit, daß die Bedeutungsgewichte jeder einzelnen Kerneigenschaft größer sind als die der Nicht-Kerneigenschaften wesentlich höher als bei der CCA. So zeigte sich in einer ersten empirischen Anwendung der CCC für Kaffeesorten, die als Pretest für die vorliegende Studie diente, daß unter Verwendung von Gleichung (3-4) zur Berechnung der Gesamtnutzenwerte für 75% der Probanden die Kerneigenschaften jeweils nicht unwichtiger waren als die wichtigste Nicht-Kerneigenschaft.[183] Ein Vergleich der Ergebnisse auf Basis Gleichung (3-3) gemäß CCA zeigte, daß bei alter Berechnungsformel der entsprechende Prozentsatz nur 14% betrug. Umgekehrt bedeutet dies, daß für 86% der Auskunftspersonen Nicht-Kerneigenschaften mindestens hinsichtlich einer Eigenschaft (teilweise um ein Vielfaches) wichtiger waren als eine oder mehrere Kerneigenschaften. Die gewählte Modifikation in Gleichung (3-4) stellt dabei nicht sicher, daß mit 100% Wahrscheinlichkeit das Bedeutungsgewicht jeder einzelnen Kerneigenschaft größer ist als das der wichtigsten Nicht-Kerneigenschaft. Darauf wurde bewußt verzichtet, denn eine solche Modellierung der Gesamtnutzenwerte würde implizieren, daß die Unterscheidung von Kern- und Nicht-Kerneigenschaften durch das Self-Explicated-Verfahren verzerrungsfrei gelingt. Davon kann jedoch vor allem in Situationen bei nur geringen Unterschieden zwischen den Self-Explicated-Bedeutungsgewichten für die Kern- und Nicht-Kerneigenschaften nicht ausgegangen werden. Welche der beiden Berechnungsformeln die tatsächlichen Präfe-

---

[183] Vgl. Hensel-Börner/Sattler (2000), S. 9.

renzen der Probanden eher abbildet, hängt maßgeblich von der Verteilung der Bedeutungsgewichte der Probanden ab und kann somit nur empirisch überprüft werden.

Um den Effekt der neuen Berechnungsformel im Vergleich zur von Srinivasan/Park (1997) vorgeschlagenen Berechnung der Gesamtnutzenwerte quantifizieren zu können, werden in der folgenden empirischen Untersuchung beide Varianten angewendet und verglichen. Zur Unterscheidung wird in diesem Zusammenhang bei Verwendung von Gleichung (3-3) von der **CCC-alt** und von der **CCC-neu** bei Berechnung der Gesamtnutzenwerte gemäß Gleichung (3-4) gesprochen. Zusammenfassend sei an dieser Stelle festgehalten, daß die CCC in ihrer Grundstruktur auf der Idee der CCA basiert, wobei zwei wesentliche Modifikationen zum Tragen kommen. Zum einen basiert die CCC auf einem computergestützten Interview, bei dem alle Daten im Rahmen eines Interviews mit der Auskunftsperson erhoben werden. Zum anderen wird vor dem Hintergrund der angeführten Überlegungen mit der CCC eine alternative Berechnungsvorschrift zur Ermittlung der Gesamtnutzenwerte vorgeschlagen. Diese zweite Modifikation wirkt sich nicht auf die Datenerhebung aus, so daß für einen vorliegenden Datensatz jeweils überprüft werden kann, welche Berechnungsformel für die Abbildung der Präferenzstrukturen am besten geeignet ist. Der Interviewablauf der CCC wird in folgender Abbildung 3-1 noch einmal im Überblick zusammengefaßt.

Abbildung 3-2: Interviewablauf der CCC

Phase	Befragungsschritt	Vorgehensweise bei der CCC
Self-Explicated Phase	1. Schritt: Eliminierung inakzeptabler Eigenschaftsausprägungen	• Direkte Frage nach inakzeptablen Eigenschaftsausprägungen (in der derzeitigen Programmversion noch nicht implementiert)
	2. Schritt: Erfassung von Präferenzen gegenüber Eigenschaftsausprägungen	• Frage nach der am wenigsten (0 Punkte) und der am meisten präferierten (10 Punkte) Ausprägung je Eigenschaft • Vergabe eines Punktwertes von 0 bis 10 für dazwischenliegende Ausprägungen
	3. Schritt: Erfassung der Wichtigkeiten	• Frage nach der wichtigsten Eigenschaft (10 Punkte) • Vergabe eines Punktwerts von 0 bis 10 Punkte für die übrigen Eigenschaften unter Verwendung der wichtigsten Eigenschaft als Ankergröße
	➔ Schätzung der Teilnutzenwerte führt zur **SE-CCC** und Identifikation der Kerneigenschaften	
Conjoint-Phase	Erfassung ganzheitlicher Präferenzen	• Angabe einer Präferenzrangfolge für 8 bis 16 Produkte, die durch 3 bis 6 Kerneigenschaften beschrieben werden • Angabe einer Kaufwahrscheinlichkeit zwischen 0% und 100% für jeden Rangplatz
Kombinationsphase	Kombination der Informationen aus den beiden vorangegangenen Befragungsteilen	• Angabe einer Präferenzrangfolge für ein Validation-Set, das aus 3 bis 12 Produkten besteht. Ebenfalls beschrieben durch 3 bis 6 Kerneigenschaften • Angabe einer Kaufwahrscheinlichkeit von 0% bis 100% für jeden Rangplatz
	➔ Ermittlung der kombinierten Gesamtnutzenwerte gemäß Gleichung (3-3) bzw. (3-4) führt zur **CCC-alt** und **CCC-neu**	

Quelle: Eigene Erstellung

3.2 Choice-Oriented Individualized Conjoint Analysis (CHIC)

Die CHIC-Variante stellt ebenfalls ein computergestütztes hybrides Verfahren dar, das analog zur CCC einen individualisierten Interviewablauf aufweist.[184] Auf Basis des kompositionellen Befragungsteils werden die individuell wichtigsten

[184] Die alternative Bezeichnung „individualized" gegenüber „customized" bei der CCC wurde lediglich vor dem Hintergrund der kreierten Abkürzung CHIC gewählt.

Eigenschaften als Kerneigenschaften identifiziert und in der folgenden Conjoint-Phase berücksichtigt. Der wesentliche Unterschied im Vergleich zur CCC liegt in der Zahl simultan zu bewertender Stimuli und darauf aufbauend in der Gestaltung der Beurteilungsaufgabe des dekompositionellen Interviewteils. In Anlehnung an die Choice-Based Conjoint Analysis werden den Probanden einzelne Choice-Sets (also Teilmengen an Stimuli) präsentiert und zunächst eine Auswahlentscheidung zur Beurteilung der Stimuli abverlangt. Darüber hinaus wird die Präferenzabfrage erweitert, um auch auf Basis der dekompositionellen Phase individuelle Teilnutzenfunktionen für jeden Probanden separat schätzen zu können. Bei der folgenden Darstellung der CHIC werden die Unterschiede gegenüber der CCC ausführlich erläutert, wogegen identische Ablaufschritte nicht noch einmal explizit beschrieben werden.

3.2.1 Self-Explicated-Phase

Der Self-Explicated-Teil der CHIC entspricht dem Vorgehen der CCC. Auf eine erneute Darstellung wird daher an dieser Stelle verzichtet und auf die Ausführungen in Kapitel 3.1.1 verwiesen.

3.2.2 Conjoint-Phase

Die dekompositionelle Phase der CHIC unterscheidet sich maßgeblich von der CCC sowohl hinsichtlich der Anzahl simultan zu bewertender Stimuli als auch des Ablaufs und Form der Präferenzabfrage. Die Konzeption der Conjoint-Phase bei der CHIC wird im folgenden vor dem Hintergrund dieser beiden Unterscheidungsmerkmale ausführlich dargestellt.

3.2.2.1 Anzahl simultan zu beurteilender Stimuli

Beschrieben werden die Stimuli der dekompositionellen Phase wie bei der CCC durch die auf Basis der Self-Explicated-Phase identifizierten Kerneigenschaften.[185] In dieser Anpassung liegt der Kern der Individualisierung des Verfahrens, denn je nach erhobener Wichtigkeit der Eigenschaften variieren die berücksichtigten Eigenschaften in der Conjoint-Phase über die verschiedenen Probanden. Während allerdings bei der CCC den Auskunftspersonen jeweils alle zu bewertenden Stimuli gleichzeitig präsentiert werden, sind diese bei der CHIC in Anleh-

---

[185] Auch hier werden wieder max. 3 Ausprägungen je Eigenschaft berücksichtigt, wobei die für jeden Probanden beste und schlechteste sowie eine mittlere Ausprägung ausgewählt werden.

nung an die Choice-Based Conjoint Analysis auf einzelne Choice-Sets aufgeteilt. Ein solches Choice-Set besteht aus einer Teilmenge der zu bewertenden Stimuli und wird jeweils unabhängig von den anderen Sets bewertet. Damit umfaßt die Entscheidung hinsichtlich der Anzahl zu bewertender Stimuli bei der CHIC zwei Fragestellungen. Zum einen muß die Zahl der insgesamt zu bewertenden Stimuli festgelegt werden und zum anderen gilt es, diese bestmöglich auf die verschiedenen Choice-Sets aufzuteilen.

Auswahl der insgesamt zu beurteilenden Stimuli

Obwohl bei der CHIC jeweils nur Teilmengen aller Stimuli simultan zu bewerten sind, können nur bei einer sehr begrenzten Anzahl an Kerneigenschaften insgesamt alle Stimuli (d.h. das vollständige Design) zur Beurteilung vorgelegt werden, um die Auskunftspersonen nicht zu überfordern. Daher muß wie bei der CCC ein reduziertes „Gesamtdesign"[186] erstellt werden. Diese Reduzierung der Stimuli erfolgt bei der CCC gemäß eines fraktionierten faktoriellen Designs. Bei der CHIC dagegen wird hierfür eine Zufallsauswahl an Stimuli aus dem jeweils vollständigen Design ausgewählt. Zwar sind zufällig erstellte bzw. randomisierte Designs in den seltensten Fällen orthogonal und büßen damit ein gewisses Maß an Effizienz[187] ein, sie ermöglichen jedoch die Schätzung von Interaktionseffekten.[188] Wie viele Stimuli insgesamt ausgewählt werden und zu bewerten sind, hängt wiederum von der Anzahl an Kerneigenschaften und den daraus resultierenden Schätzparametern ab.

Über eine zufällige Auswahl hinaus wird die Reduzierung der Stimuli an dem Prinzip der Pareto-Optimalität orientiert. Ein pareto-optimales Design ist dadurch gekennzeichnet, daß keine Alternative alle anderen dominiert. Die Pareto-Optimalität wird dabei im Sinne von Krieger/Green (1991) definiert. Entsprechend dieser Definition dominiert eine Alternative eine andere, sofern sie in wenigstens einer Eigenschaftsausprägung eindeutig präferiert wird, während alle

---

186 Das Gesamtdesign bezieht sich auf alle innerhalb des ganzen Conjoint-Interviews zu bewertenden Stimuli und nicht auf die Anzahl an Stimuli innerhalb eines Choice-Sets.

187 Ein Experimentdesign gilt als effizient, wenn die Varianzen und Kovarianzen der Parameterschätzungen minimiert werden (vgl. Kuhfeld/Tobias/Garrat (1994), S. 545).

188 Im Gegensatz dazu können insbesondere bei orthogonalen Designs lediglich die Haupteffekte geschätzt werden. Einzelne Interaktionseffekte können nur dann ermittelt werden, sofern diese explizit im Experimentdesign integriert werden (vgl. Green/Srinivasan (1978), S. 110).

anderen Eigenschaften bei beiden Alternativen in der gleichen Ausprägung auftreten.[189] Der wesentliche Vorteil eines solchen Designs liegt darin, daß von den Auskunftspersonen nur solche Entscheidungen erfragt werden, deren Antwort, auf Basis bereits gewonnener Informationen, nicht a priori festzulegen bzw. zu erwarten ist. Unterscheiden sich beispielsweise zwei zu beurteilende Kühlschränke nur in den Ausprägungen des Energieverbrauchs, kann davon ausgegangen werden, daß der Kühlschrank mit niedrigerem Verbrauch von den Probanden präferiert wird. Damit wird der Informationsgehalt des gesamten Interviews maximiert, indem mit jeder Entscheidung der Probanden möglichst neue Informationen erhoben werden.[190]

Zur konkreten Umsetzung dieser Bedingung werden bei der CHIC für den Conjoint-Teil nur die Eigenschaftskombinationen aus dem vollständigen Design ausgewählt, die bei mindestens zwei Eigenschaften unterschiedliche Ausprägungen aufweisen. Damit wird die Wahrscheinlichkeit, daß dominante Alternativen in einem Choice-Set auftreten, maßgeblich verringert, jedoch nicht vollkommen ausgeschlossen. Auf eine Sicherstellung der Pareto-Optimalität aller Stimuli wurde verzichtet, da eine Konstruktion von Choice-Sets mit mehr als 2 Stimuli kaum noch möglich ist, bei denen nicht zumindest paarweise Dominanz auftritt. Darüber hinaus liefern die Antworten hinsichtlich dominierter Alternativen eine Möglichkeit, die Reliabilität, d.h. die Konsistenz des Antwortverhaltens der Probanden zu überprüfen.

Aufteilung der Stimuli auf die Choice-Sets

Die zweite Fragestellung hinsichtlich der Designkonstruktion umfaßt wie bei allen Choice-Experimenten die Aufteilung der Stimuli auf die einzelnen Choice-Sets. Eine Möglichkeit besteht darin, entsprechend fest vorgegebener Designplä-

---

189 Vgl. Krieger/Green (1991), S. 337.

190 Es sei an dieser Stelle angemerkt, daß pareto-optimale Designs nicht notwendigerweise orthogonal sind. Vielmehr stehen diese beiden Zielkriterien teilweise im Konflikt zueinander, schließen sich aber nicht grundsätzlich aus. Krieger/Green (1991) schlagen beispielsweise eine Heuristik zur Designkonstruktion vor, bei der ausgehend von einem orthogonalen Design, die Pareto-Optimalität maximiert wird. Diese Heuristik kann zu gleichzeitig orthogonalen und pareto-optimalen Designs führen, stellt dies aber nicht in jedem Fall sicher (vgl. Krieger/Green (1991), S. 343). Eine Verletzung der Orthogonalität bedeutet inhaltlich, daß zwischen den Eigenschaften des Designs Multikollinearität besteht, diese also nicht unabhängig voneinander sind. Inwieweit sich diese Prämissenverletzung auf die Schätzergebnisse auswirkt, kann a priori nicht getestet werden, sondern wird nachträglich bei der Datenauswertung überprüft.

ne vorzugehen.[191] Alternativ dazu können die Stimuli auch zufällig auf die verschiedenen Choice-Sets aufgeteilt werden.[192] Die CHIC als computergestütztes Interview ermöglicht eine solch zufällige Konstruktion der Choice-Sets und wird in der vorliegenden Programmversion angewandt. Damit wird neben der Eigenschaftsauswahl für die Conjoint-Phase auch die Zusammenstellung der Choice-Sets individualisiert. Selbst wenn zwei Probanden gleiche Kerneigenschaften aufweisen und identische Präferenzen gegenüber den Ausprägungen geäußert haben, müssen die konstruierten Choice-Sets nicht übereinstimmen. Mit der zufälligen Aufteilung wird das Ausmaß der Individualisierung des Interviews zusätzlich erhöht und die Gefahr systematischer Verzerrungen aufgrund der Designerstellung verringert.

Bisher wurde erläutert, wie die Stimuli auf die verschiedenen Choice-Sets aufgeteilt werden können. In einem weiteren Schritt der Designkonstruktion sind deren Größe und Anzahl zu definieren. Zu berücksichtigen ist dabei, daß diese beiden Entscheidungen nicht unabhängig voneinander getroffen werden können. Bei gegebener Stimulizahl hängt die Anzahl zu konstruierender Choice-Sets unmittelbar davon ab, wie viele Alternativen je Set betrachtet werden. Es gilt, einen Trade-Off zu machen zwischen wenigen umfangreichen einerseits und mehreren kleinen Choice-Sets andererseits. Die Frage nach der optimalen Anzahl an Alternativen in einem Choice-Set ist in der Literatur bisher kaum bearbeitet worden. So werten Johnson/Orme insgesamt 21 kommerzielle Choice-Based Conjoint-Studien aus und untersuchen u.a. den Einfluß der Anzahl an Choice-Sets insgesamt, geben aber keine Auskunft darüber, wie groß die Sets in den Studien jeweils waren.[193]

In der vorliegenden Version der CHIC werden in jedem Choice-Set 4 Produktkonzepte und zusätzlich die Nicht-Kauf-Alternative berücksichtigt.[194] Mit der

---

191 Zu den Grundlagen der Designerstellung bei Choice-Experimenten vgl. u.a. Louviere/Woodworth (1983), Kuhfeld/Tobias/Garratt (1994) und Huber/Zwerina (1996).

192 Auch Louviere/Woodworth (1983) schlagen eine zufällige Designgestaltung als eine alternative Vorgehensweise vor (vgl. Louviere/Woodworth (1983), S. 354). Die Choice-Based Conjoint Analysis von Sawtooth Software bietet ebenfalls die Möglichkeit einer zufälligen Designerstellung (vgl. Sawtooth Software (1995), S. 4).

193 Vgl. hierzu Johnson/Orme (1996).

194 Damit liegt die Anzahl simultan zu bewertender Stimuli bei der CHIC mit 4+1 Alternativen zwischen den anderen betrachteten Verfahren CCC (9 Stimuli) und ACA (2 Stimuli).

Nicht-Kauf-Alternative haben die Probanden die Möglichkeit, alle Produktalternativen eines Choice-Sets abzulehnen, sofern keine für sie attraktive Alternative angeboten wird. Mit dieser Größe der Choice-Sets werden bewußt mehr Stimuli pro Set präsentiert als bei der ACA, deren dekompositioneller Teil auf Paarvergleichen basiert. Die Konstruktion größerer Choice-Sets ist prinzipiell möglich, in den meisten Untersuchungen jedoch nicht erforderlich, da die Gesamtzahl zu bewertender Stimuli durch die (nahezu) pareto-optimale Designkonstruktion stark reduziert wird. So unterscheiden sich beispielsweise in einem 3*3*3*2 –Design[195] aus 4 Kerneigenschaften lediglich 14 Stimuli in mindestens 2 Ausprägungen und erfüllen somit die zur Designkonstruktion formulierte Nebenbedingung. Bei gegebener Anzahl an Alternativen insgesamt und pro Choice-Set ist die Zahl zu konstruierender Choice-Sets determiniert. Je nach Anzahl an Kerneigenschaften und deren Ausprägungen werden bei der aktuellen Version der CHIC lediglich 3 oder 4 Choice-Sets konstruiert. Diese Größenordnung ist nicht mit den Designs bei klassischen Choice-Experimenten zu vergleichen, bei denen die Probanden meist 16 bis 20 oder mehr Choice-Sets bewerten müssen.[196] Dieser deutliche Unterschied zwischen der CHIC und der eigentlichen Choice-Based Conjoint Analysis begründet sich darin, daß bei der CHIC für jedes Choice-Set wesentlich mehr Informationen erhoben werden als lediglich die einmalige Auswahlentscheidung der am meisten präferierten Alternative, wie das bei der CBC der Fall ist. Die Präferenzabfrage der CHIC wird im folgenden Kapitel ausführlich erläutert.

3.2.2.2 Form der Beurteilungsaufgabe

Die Beurteilungsaufgabe der CHIC wird in ihrer Form zunächst an der klassischen Choice-Based Conjoint Analysis orientiert. Durch eine Auswahlentscheidung, wie sie bei der Choice-Based Conjoint Analysis von den Probanden abverlangt wird, soll insbesondere der Realitätsbezug der Aufgabenstellung erhöht werden. Mit der CHIC wird jedoch zudem das Ziel verfolgt, individuelle Teilnut-

---

195 Dieses Design impliziert, daß drei Kerneigenschaften 3 Ausprägungen und eine Kerneigenschaft lediglich 2 Ausprägungen aufweisen. Das vollständige Design umfaßt demnach insgesamt 54 Eigenschaftskombinationen.

196 Vgl. Elrod/Louviere/Davey (1992), S. 371; Louviere/Gaeth (1988), S. 68; Louviere/Fox/Moore (1993), S. 206; Louviere/Woodworth (1983), S. 355 ff.; Moore/Gray-Lee/Louviere (1998), S. 200; Oliphant et al. (1992), S. 386. Die Programmversion von Sawtooth Software ermöglicht bei der Choice-Based Conjoint Analysis sogar bis zu 50 Abfragen.

zenwerte auch auf Basis der dekompositionellen Interviewphase separat bestimmen zu können. Die klassische Choice-Based Conjoint Analysis ermöglicht dagegen nur eine aggregierte Auswertung. Vor diesem Hintergrund wird die Datenerhebung der CHIC erweitert und umfaßt in der dekompositionellen Phase folgende drei Teilschritte, die für jedes Choice-Set wiederholt werden:

1. Schritt: Auswahlentscheidung

Der erste Schritt der Beurteilungsaufgabe entspricht dem Vorgehen der Choice-Based Conjoint Analysis und umfaßt eine Auswahlentscheidung aus dem präsentierten Choice-Set. Die Probanden werden gebeten, das Produkt auszuwählen, welches sie am ehesten kaufen bzw. für welches sie sich in der entsprechenden Kaufsituation entscheiden würden. Dabei haben die Auskunftspersonen auch die Möglichkeit, sich für keine der zur Wahl stehenden Alternativen zu entscheiden. Hierfür wird wie bei der Choice-Based Conjoint Analysis in jedem Choice-Set zusätzlich eine Nicht-Kauf-Alternative berücksichtigt. Die Integration dieser Nicht-Kauf-Alternative stellt einen wesentlichen Vorteil der Choice-Based Conjoint Analysis dar, da mit ihr zusätzlich der Realitätsbezug der Aufgabenstellung erhöht wird.[197]

Ein mit der klassischen Choice-Based Conjoint Analysis verbundener Nachteil betrifft die Ineffizienz der skizzierten Datenerhebung. Während der Proband lediglich die für ihn beste Alternative pro Choice-Set identifiziert, müssen für diese Entscheidung die Informationen hinsichtlich aller Alternativen des Sets verarbeitet werden. Dieses Ungleichgewicht wird auch darin deutlich, daß der Erhebungsaufwand durch die zusätzliche Angabe der zweit- oder drittbesten Alternative nur unwesentlich steigt, der zusätzliche Informationsgewinn aber im Verhältnis dazu wesentlich höher ist.[198] Um die Effizienz der Datenerhebung zu erhöhen und die Datengrundlage für eine individuelle Schätzung der Teilnutzenwerte zu schaffen, wird die Auswahlentscheidung auch innerhalb der Choice-Sets wiederholt. Das Choice-Set wird hierfür um die am meisten präferierte Alternative reduziert und die Probanden erneut befragt, welche Alternative sie aus diesem verkleinerten Set am ehesten auswählen würden. Nach Abschluß dieser wiederholten Auswahlentscheidung liegt damit implizit eine Rangfolge mit absteigender

---

[197] Vgl. hierzu auch Oliphant et al. (1992), S. 385; Louviere/Gaeth (1988), S. 62.

[198] Vgl. Johnson/Orme (1996), S. 11.

Präferenz für die Produkte innerhalb eines Choice-Sets vor. Sofern die Nicht-Kauf-Alternative gewählt wird, werden im Anschluß dennoch die Präferenzen zwischen den Alternativen des Choice-Sets erfragt. Die Bewertung erfolgt dann unter der Annahme, daß keine anderen Produktalternativen zur Verfügung stehen. Somit werden zumindest noch Informationen zu den Präferenzen zwischen den verschiedenen Alternativen innerhalb des betreffenden Sets erhoben.[199] Die Wahl der Nicht-Kauf-Alternative wird anschließend bei der Ableitung einer Gesamtrangfolge für alle bewerteten Stimuli berücksichtigt.

2. Schritt: Quantifizierung der Präferenzunterschiede

Nach den einzelnen Auswahlentscheidungen liegt implizit eine Präferenzrangfolge für die Alternativen innerhalb eines Choice-Sets vor. Ziel des zweiten Befragungsschritts ist es, die Präferenzunterschiede zwischen den ermittelten Rangplätzen zu quantifizieren. Dies liefert detaillierte Informationen hinsichtlich der Präferenzen und ist zudem Voraussetzung für eine Vergleichbarkeit der Präferenzurteile zwischen den verschiedenen Choice-Sets. Um diese Bewertung für die Probanden zu vereinfachen, werden die Produktalternativen auf dem Bildschirm gemäß der implizit erhobenen Rangfolge sortiert. Die Quantifizierung der Präferenzunterschiede wird mit einer von 0 bis 100 Punkten definierten Rating-Skala vorgenommen, wobei ein höherer Wert eine höhere Präferenz bedeutet. Die Endpole dieser Rating-Skala werden wiederum mit der am meisten präferierten (100 Punkte) und am wenigsten präferierten (0 Punkte) Alternative fixiert und dienen als Ankerpunkte bei der Bewertung der anderen Alternativen.[200] Aufgabe der Probanden ist dann, den mittleren Alternativen entsprechend der individuellen Präferenz einen Wert zwischen 0 und 100 Punkten zuzuweisen. Die Rating-Skala wird gleichzeitig auf dem Bildschirm grafisch abgebildet, um die Unterschiede zwischen den bewerteten Präferenzen sowohl anhand der zugeordneten Punktwerte als auch grafisch zu visualisieren. Damit wird die Gefahr eines „information overload", d.h. einer Informationsüberlastung der Probanden reduziert.[201]

---

199 Vgl. hierzu auch Elrod/Louviere/Davey (1992), S. 370.

200 Für diese Verankerung spricht, wie bei der Erfassung der Wichtigkeiten für die Eigenschaften in der Self-Explicated-Phase, daß dadurch die Wahrscheinlichkeit erhöht wird, die Präferenzunterschiede tatsächlich auf einer Ratio-Skala zu messen.

201 Vgl. Loosschilder et al. (1995), S. 20.

3. Schritt: Transformation der Präferenzunterschiede in eine monetäre Größe

Auf Basis der Rating-Skala werden die Unterschiede innerhalb eines Choice-Sets quantifiziert. Um die Vergleichbarkeit der Präferenzunterschiede zwischen den einzelnen Choice-Sets zu gewährleisten, muß sichergestellt werden, daß die Bewertungsvorschrift von jedem Proband in jedem Set gleich interpretiert wird. Durch die Verankerung der am meisten präferierten und am wenigsten präferierten Alternative werden dazwischenliegende Produktkonzepte immer in Relation dazu auf der Rating-Skala bewertet werden. Problematisch allerdings ist, daß dadurch die Bewertung eines Produkts auf dieser Skala immer davon abhängt, welche Alternativen innerhalb des jeweiligen Choice-Sets noch zur Wahl stehen.[202] So kann beispielsweise eine Alternative aus einem Choice-Sets als Erstpräferenz hervorgehen und mit 100 Punkten bewertet werden, die innerhalb eines anderen Choice-Sets als wesentlich schlechter beurteilt würde.

Eine individuelle Schätzung der Teilnutzenwerte auf Basis der dekompositionellen Phase, wie sie bei der CHIC beabsichtigt ist, setzt jedoch gerade die Vergleichbarkeit der Präferenzurteile zwischen den einzelnen Choice-Sets voraus. Um dem Effekt einer unterschiedlichen Skaleninterpretation und damit eingeschränkter Vergleichbarkeit zwischen verschiedenen Choice-Sets entgegenzuwirken, werden die Präferenzunterschiede in eine monetäre Größe transformiert. Dafür wird die Auskunftsperson in diesem dritten Schritt der Stimulibewertung gebeten, den Unterschied zwischen zwei Stimuli des Choice-Sets in einem DM-Betrag zu bewerten. Damit können die im zweiten Schritt der Beurteilungsaufgabe erhobenen Punktwerte in DM-Beträge umgerechnet und Unterschiede in der Interpretation der Punktwerte der Rating-Skala bereinigt werden. Neben einer besseren Vergleichbarkeit der Skalenwerte über die Choice-Sets bietet diese Transformation der Skalenwerte in DM-Beträge zusätzlich die Möglichkeit, die Präferenzunterschiede in Form einer direkt angegebenen Zahlungsbereitschaft für die entsprechenden Eigenschaftsausprägungen zu messen. Nach allen 3 beschrie-

---

202 Diese Problematik wird in der Literatur unter dem Begriff „Independence from Irrelevant Alternatives (kurz: IIA)“ diskutiert und besagt, daß der Gesamtnutzen einer Alternative unabhängig davon ist, welche Alternativen noch im Choice-Set vorhanden sind. Dies Annahme stellt ebenfalls eine Besonderheit der klassischen Choice-Based Conjoint Analysis dar. Zur Definiton der IIA-Eigenschaft vgl. Louviere/Woodworth (1983), S. 352.

benen Befragungsschritten liegt für die Stimuli *innerhalb* eines Choice-Sets eine Präferenzrangfolge vor, deren Unterschiede in DM-Beträgen quantifiziert sind.

3.2.2.3 Schätzung der Teilnutzenwerte

Eine individuelle Auswertung der Präferenzdaten beispielsweise mittels einer OLS-Schätzung ist erst dann möglich, wenn als abhängige Variable einer solchen Schätzung eine Gesamtreihenfolge aller bewerteten Stimuli vorliegt. Bisher liegen metrische Präferenzaussagen nur hinsichtlich der Stimuli innerhalb der einzelnen Choice-Sets vor. Um Stimuli aus unterschiedlichen Choice-Sets miteinander vergleichen zu können und alle bewerteten Produkte in eine bewertete Gesamtreihenfolge sortieren zu können, müssen die Ergebnisse der Choice-Sets miteinander verankert werden. Eine Möglichkeit hierfür besteht darin, bereits bei der Konstruktion der Choice-Sets eine entsprechende Verankerung sicherzustellen. Treten in jedem Choice-Sets wenigstens 2 Stimuli immer wieder auf, ist ein direkter Vergleich der Skalenwerte zwischen der Sets möglich. Die in allen Sets enthaltenen Stimuli bilden eine „Brücke" zwischen den Sets und dienen als Ankerpunkt für alle anderen Vergleiche. Wenn in jedem Choice-Set der Präferenzunterschied zwischen zwei identischen Stimuli quantifiziert ist, werden die Skalenwerte über alle Choice-Sets miteinander vergleichbar. Gleichzeitig ist damit die Unabhängigkeit der Skaleninterpretation von der Zusammensetzung eines Sets sichergestellt. Nachteilig an einer solchen Designkonstruktion ist allerdings die damit verbundene Erhöhung des Erhebungsaufwands. Pro Choice-Set werden jeweils nur Informationen für 2 weitere Stimuli erhoben. Bei konstanter Stimulizahl sind folglich insgesamt mehr Choice-Sets zu bewerten. Eine weitere Gefahr ist, daß Probanden die wiederholte Bewertung identischer Stimuli als Konsistenzprüfung auffassen und anstatt tatsächlicher Präferenzunterschiede bewußt immer wieder die gleichen Skalenwerte angeben.

Für die derzeitige Programmversion der CHIC wird daher eine alternative Vorgehensweise gewählt, um die bewerteten Stimuli in eine metrische Gesamtreihenfolge zu bringen. Die Verankerung der einzelnen Choice-Sets wird nachträglich auf Basis der Informationen aus der Self-Explicated-Phase vorgenommen. Mit Hilfe der Self-Explicated-Teilnutzenwerte für die Eigenschaft „Preis" werden die Gesamtnutzenwerte der verschiedenen Stimuli in DM-Beträge transformiert, die als maximale Zahlungsbereitschaft für das jeweilige Produkt interpretiert werden

kann.[203] Diese Zahlungsbereitschaft stellt dabei eine absolute Größe dar, die einen direkten Vergleich über die verschiedenen Sets ermöglicht. Berechnet man für alle Stimuli auf Basis der Self-Explicated-Teilnutzen die jeweilige Zahlungsbereitschaft, liegt damit eine metrische Gesamtreihenfolge aller Stimuli vor. Dies würde allerdings implizieren, daß die Erfassung der Präferenzstrukturen auf Basis der Self-Explicated-Phase verzerrungsfrei gelingt. Da davon nicht ausgegangen werden kann, wird lediglich die Zahlungsbereitschaft für den Stimulus in jedem Choice-Set berechnet, der anschließend zur Verankerung der Skalen zwischen den Choice-Sets dient. In der vorliegenden Untersuchung wird die Zahlungsbereitschaft für das in jedem Choice-Set am schlechtesten bewertete Produkt berechnet.[204]. Damit werden die Endpole der Rating-Skala miteinander verankert und sind direkt vergleichbar. Die durch die Schritte 1 bis 3 des Conjoint-Interviews ermittelten DM-Beträge für jedes Choice-Set werden daraufhin so umskaliert, daß zum einen die Präferenzunterschiede innerhalb eines Choice-Sets abgebildet werden und zum anderen die Choice-Sets als ganzes entsprechend der Self-Explicated-Zahlungsbreitschaften miteinander verankert werden. Die Vorgehensweise zur Verankerung der verschiedenen Choice-Sets wird anhand eines vereinfachten Beispiels in Abbildung 3-3 mit 2 Sets veranschaulicht:

---

203 Einschränkend muß angemerkt werden, daß die maximale Zahlungsbereitschaft nicht explizit als solche von den Probanden erhoben wurde, sondern aus den gegebenen Informationen abgeleitet wird. Da Abweichungen zur tatsächlich maximalen Zahlungsbereitschaft der Probanden auftreten können, ist die vorgenommene Interpretation nicht unproblematisch. Für die Vergleichbarkeit der Präferenzunterschiede zwischen den alternativen Produktalternativen ist die Erfassung der tatsächlich maximalen Zahlungsbereitschaft nicht zwingend erforderlich, was die Problematik der vorgenommenen Interpretation zumindest teilweise reduziert.

204 Alternativ können sowohl die jeweils am besten oder dazwischen eingestufte Produkte zur Verankerung herangezogen werden. Empfohlen wird aber eine der beiden Eckwerte der Skala zu verwenden, da sich in der vorliegenden empirischen Untersuchung zeigte, daß sich dabei die geringsten Inkonsistenzen zwischen der auf Basis der Self-Explicated-Befragung geschätzten und im Conjoint-Teil beobachteten Reihenfolge ergab.

Abbildung 3-3: Beispiel zur Umskalierung der Präferenzwerte in der Conjoint-Phase der CHIC

a) Vorliegende Bewertungen innerhalb eines Choice-Sets nach der Datenerhebung:
(Punktewerte der Rating-Skala sind bereits in DM-Beträge transformiert)

Choice-Set 1	
Stimulus A	650,- DM
Stimulus B	400,- DM
Stimulus C	180,- DM
Stimulus D	0,- DM

Choice-Set 2	
Stimulus E	400,- DM
Stimulus F	200,- DM
Stimulus G	100,- DM
Stimulus H	0,- DM

b) Auf Basis der Self-Explicated-Phase ermittelte maximale Zahlungsbereitschaften für die beiden am wenigsten präferierten Stimuli je Choice-Set:

Stimulus D	20,- DM

Stimulus H	200,- DM

c) Umskalierte Präferenzwerte nach Verankerung der beiden Choice-Sets:

Choice-Set 1	
Stimulus A	650,- DM
Stimulus B	400,- DM
Stimulus C	180,- DM
Stimulus D	0,- DM

Choice-Set 2	
Stimulus E	580,- DM
Stimulus F	380,- DM
Stimulus G **	280,- DM
Stimulus H *	180,- DM

Anmerkung:
* Der Präferenzwert für Stimulus H beträgt jetzt 180,- DM, da die Zahlungsbereitschaft (nach SE-Daten) um diesen Betrag höher ist als die des am wenigsten präferierten Stimulus D aus Set 1
=> Verankerung der Choice-Sets zueinander

** Die Bewertungen der Conjoint-Phase haben ergeben, daß der Präferenzwert für Stimuli G um 100,- DM höher gegenüber Stimulus H ist.
=> Berücksichtigung der Präferenzen innerhalb eines Choice-Sets ermittelt aus der Conjoint-Phase des Interviews.

Quelle: Eigene Erstellung

Nach dieser Umskalierung der DM-Beträge können alle bewerteten Stimuli in eine metrische Gesamtreihenfolge gebracht werden. Diese dient als Grundlage für eine Schätzung der Teilnutzenwerte für die Kerneigenschaften. Aufgrund der vorliegenden metrischen Skalenniveaus empfiehlt sich hier die Anwendung einer

OLS-Schätzung, um alle vorhandenen Informationen auch in der Schätzung zu berücksichtigen.[205]

3.2.3 Kombinationsphase

Wie bei der CCC liegen bei der CHIC für die Kerneigenschaften des i-ten Probanden C_i sowohl Teilnutzenwerte unter Verwendung der Informationen der Self-Explicated- als auch der Conjoint-Phase vor. Grundidee der CHIC ist auch hier, diese beiden Teilnutzenwerte optimal zu kombinieren, um die tatsächlichen Präferenzstrukturen der Probanden bestmöglich abzubilden. Als Referenzgröße dienen analog zur CCC, die erhobenen Präferenzen für ein zusätzliches Validation-Set. Ziel ist es, diese beobachteten Präferenzurteile auf Basis der kombinierten Teilnutzenwerte aus Self-Explicated- und Conjoint-Phase möglichst gut zu prognostizieren. Als Validation-Set werden die Präferenzen eines zusätzlichen Choice-Sets herangezogen, welche nicht in die Schätzung der Conjoint-Teilnutzenwerte eingehen. Die Vorgehensweise zur Ermittlung der optimalen Kombination der Teilnutzenwerte entspricht dem der CCC auf deren wiederholte Darstellung an dieser Stelle verzichtet wird. Zudem werden ebenfalls aufgrund der in Kapitel 3.1.3 angeführten Argumentation die beiden alternativen Berechnungsformeln zur Bestimmung der Gesamtnutzenwerte bei neuen Produkten herangezogen. Zur Unterscheidung der jeweils eingesetzten Schätzmethode wird im folgenden in Anlehnung an die Notation der CCC zwischen CHIC-alt und CHIC-neu differenziert. Abbildung 3-4 faßt den Interviewablauf der CHIC noch einmal zusammen.

---

[205] Vgl. Green/Srinivasan (1978), S. 113.

Abbildung 3-4: Interviewablauf der CHIC

Phase	Befragungsschritt	Vorgehensweise bei der CHIC
Self-Explicated-Phase	Zur Vorgehensweise vgl. Self-Explicated-Phase der CCC (vgl. Abbildung 3-2)	
	➔ Ermittlung der Teilnutzenwerte führt zur **SE-CHIC**	
Conjoint-Phase je Choice-Set	1. Schritt: Auswahlentscheidungen ➢ Präferenzrangfolge innerhalb eines Choice-Sets wird erfaßt	• Auswahlentscheidung aus einem Choice-Set bestehend aus 4 bis 6 Alternativen, die durch die 3 bis 6 wichtigsten Eigenschaften beschrieben sind • Wiederholung der Auswahlentscheidung aus den verbliebenen Alternativen
	2. Schritt: Quantifizierung der Präferenzunterschiede ➢ Präferenzunterschiede innerhalb eines Choice-Sets werden quantifiziert	• Auf einer Rating-Skala mit 100 bis 0 Punkten werden die zuerst ausgewählte (100 Punkte) und die zuletzt ausgewählte Alternative (0 Punkte) fixiert • Vergabe von 0 bis 100 Punkten für dazwischenliegende Alternativen die an zweiter, dritter usw. Stelle ausgewählt wurden. Die beiden Extrempole dienen dabei als Ankerpunkte
	3. Schritt: Transformation der Präferenzunterschiede in eine monetäre Größe ➢ Präferenzunterschiede werden in DM ausgedrückt und Vergleichbarkeit über verschiedene Sets wird gewährleistet	• Direkte Frage nach einem DM-Betrag zur Quantifizierung des Präferenzunterschieds zwischen 2 Alternativen des Choice-Sets
Kombinationsphase	Kombination der Informationen aus den beiden vorangegangenen Befragungsteilen	• Wiederholung der 3 Befragungsschritte der Conjoint-Phase für mindestens ein zusätzliches Choice-Set
	➔ Ermittlung der kombinierten Gesamtnutzenwerte gemäß Gleichung (3-3) bzw. (3-4) führt zur **CHIC-alt** bzw. **CHIC-neu**	

Quelle: Eigene Erstellung

3.3 Adaptive Conjoint Analysis (ACA)

Die von Johnson (1987) entwickelte ACA hat sich mittlerweile insbesondere in der kommerziellen Anwendung zum Standardinstrument für die Messung von Nachfragerpräferenzen etabliert. Sie stellt das bisher einzig computergestützte adaptive Verfahren dar, das als kommerzielle Software erhältlich ist.[206] Ein wesentlicher Vorteil der ACA gegenüber den beiden bisher vorgestellten Verfahren

[206] Vgl. Carmone/Schaffer (1995), S. 117.

CCC und CHIC liegt darin, daß die gesamte Datenerhebung computergestützt abläuft und gleichzeitig die Datenauswertung bis hin zu verschiedenen Simulationsmodellen in der Software integriert ist. Diese Benutzerfreundlichkeit in der Anwendung ist sicherlich ein Grund, für die sehr weite Verbreitung des Verfahrens.[207] Die ACA wird als drittes Verfahren der computergestützten hybriden Ansätze dargestellt. Während der Ablauf der ACA von anderen Autoren in 4 verschiedene Phasen eingeteilt wird, werden die folgenden Ausführungen aus Gründen der Vergleichbarkeit zur CCC und CHIC entsprechend der bisherigen Vorgehensweise gegliedert.[208] Gleichzeitig werden die Unterschiede zu den beiden anderen Verfahren bei den einzelnen Teilschritten aufgezeigt.

3.3.1 Self-Explicated-Phase

Die Datenerhebung des kompositionellen Interviewteils der ACA umfaßt wie CCC bzw. CHIC 3 Teilschritte, die sich in ihrer Gestaltung aber teilweise von denen der SE-CCC bzw. SE-CHIC unterscheiden.

3.3.1.1 Datenerhebung

1. Schritt: Eliminierung inakzeptabler Eigenschaftsausprägungen

Dieser erste Teilschritt ermöglicht der Auskunftsperson jeweils inakzeptable Ausprägungen bei einer Eigenschaft zu eliminieren und vom folgenden Interviewablauf auszuschließen. Es handelt sich dabei um einen Befragungsschritt, der vom Marktforscher optional in die Untersuchung integriert werden kann.[209]

2. Schritt: Erfassung von Präferenzen zu den Eigenschaftsausprägungen

Zur Beurteilung der Präferenzen hinsichtlich der Eigenschaftsausprägungen kann der Anwender bei der ACA zwischen zwei alternativen Vorgehensweisen wählen. Zum einen besteht die Möglichkeit, Präferenzrangfolgen für die Ausprägungen jeder Eigenschaft zu erfragen. Zum anderen kann eine vom Benutzer definierte Rating-Skala (mit 2 bis 9 Kategorien) eingesetzt werden, um Präferenzen gegenüber den verschiedenen Eigenschaftsausprägungen zu erheben.[210]

---

207 Vgl. Green/Srinivasan (1990), S. 11.

208 Vgl. Green/Krieger/Agarwal (1991), S. 216; Reiners (1996), S. 115 Weiber/Rosendahl (1997), S. 110.

209 Vgl. Johnson (1987), S. 259.

210 Vgl. Metegrano (1994), S. 3-4.

Die vorliegende Software der ACA ermöglicht es, bis zu 9 Ausprägungen einer Eigenschaft in einer Studie zu integrieren. Trotzdem wird wie bei der CCC und der CHIC nur eine Teilmenge dieser Ausprägungen in der folgenden dekompositionellen Phase berücksichtigt. Während bei der CCC und CHIC max. 3 Ausprägungen ausgewählt werden, liegt die Obergrenze bei der ACA bei 5 Ausprägungen. Die Auswahl kann direkt durch das System erfolgen. Berücksichtigt werden dann die Eigenschaftsausprägungen, die im vorangegangenen Befragungsschritt am meisten präferiert wurden. Alternativ können die Probanden nach den 5 Ausprägungen gefragt werden, deren Auftreten bei einem gewählten Produkt am wahrscheinlichsten ist. Letzteres Vorgehen bietet den Vorteil zwischen präferierten und tatsächlich akzeptierten Ausprägungen zu differenzieren. Abweichungen zwischen präferierten und tatsächlich akzeptierten Ausprägungen sind durchaus denkbar. So können Probanden eine bestimmte Marke durchaus präferieren, würden sich aber z.B. unter Berücksichtigung des Preises dieser Marke nicht für diese entscheiden.[211]

Auch die ACA gibt dem Marktforscher die Möglichkeit, auf die Präferenzabfrage bei quantitativen Eigenschaften zu verzichten.[212] Unter der Annahme, daß alle anderen Eigenschaften konstant sind, kann beispielsweise die Frage nach der Präferenz zwischen verschiedenen Preisen in den meisten Untersuchungen ausgeklammert werden. Durch den Verzicht auf solche a priori festzulegende Abfragen kann die Interviewzeit verkürzt und die Effizienz der Datenerhebung erhöht werden. Im Gegensatz zur CCC und CHIC entfällt bei der ACA in diesem Fall die Bewertung der Ausprägungen dieser Eigenschaft insgesamt.[213] Damit wird bei der ACA implizit für diese Eigenschaften eine über alle Ausprägungen lineare Teilnutzenfunktion unterstellt. Diese einschränkende Annahme ist durchaus kritisch zu bewerten, denn sie impliziert z.B. für die Eigenschaft „Preis", daß jede relative Preisänderung sich in gleichem Maße auswirkt. Davon kann jedoch nicht grundsätzlich ausgegangen werden. Vor allem beim Überschreiten möglicher

---

211 Vgl. Metegrano (1994), S. 3-5 f.

212 Vgl. Johnson (1987), S. 260.

213 Bei den beiden anderen Verfahren werden an dieser Stelle lediglich die jeweils am meisten und am wenigsten präferierte Ausprägung a priori festgelegt. Alle weiteren Ausprägungen sind von den Probanden zwischen diesen Endpunkten einzuordnen.

Preisschwellen ist kein linearer Zusammenhang zwischen dem Preis und dem damit verbundenen Nutzen für die Probanden zu erwarten.[214]

3. Schritt: Abfrage der Wichtigkeiten der Eigenschaften

Nachdem die für jede Testperson max. 5 relevanten Ausprägungen identifiziert sind, werden die Wichtigkeiten der Eigenschaften erfragt. Hierfür erscheinen die jeweils beste und schlechteste Ausprägung einer Eigenschaft. Damit wird auch bei diesem Verfahren die Frage nach der Wichtigkeit einer Eigenschaft in Verbindung mit deren Ausprägungen gestellt. Die Bedeutung des Unterschieds zwischen bester und schlechtester Eigenschaftsausprägung, als Definition für die Wichtigkeit einer Eigenschaft, wird wiederum auf einer Rating-Skala bewertet, für die zwischen 2- und 9-Skalenpunkte festgelegt werden können.[215]

In diesem Befragungsschritt liegt der bedeutendste Unterschied zwischen dem Self-Explicated-Teil der ACA und dem der beiden neuen Verfahren. Während bei der CCC und der CHIC die jeweils wichtigste Eigenschaft für die Beurteilung der anderen Eigenschaften als Ankergröße fungiert, werden bei der ACA alle Eigenschaften separat betrachtet. Für jede Eigenschaft wird jeweils unabhängig von allen anderen die Frage nach der Wichtigkeit wiederholt. Diese separate Beurteilung birgt die Gefahr, daß die Skalen nicht über alle Eigenschaften hinweg konsistent interpretiert werden und deren direkter Vergleich dadurch nur eingeschränkt zulässig ist. Darüber hinaus verlangt die Self-Explicated-Phase der CCC und der CHIC bereits an dieser Stelle des Interviews eine relative Bewertung der Eigenschaften zueinander. Dieses erste Abwägen erhöht gleichzeitig die Wahrscheinlichkeit, die Wichtigkeiten der Eigenschaften tatsächlich auf einer Ratio-Skala zu erheben, was Voraussetzung für deren metrische Interpretation ist.[216]

3.3.1.2 Schätzung der Teilnutzenwerte

Die ACA berechnet ebenfalls bereits auf Basis der in der Self-Explicated-Phase erhobenen Daten Teilnutzenwerte für die Eigenschaftsausprägungen. Die Schätzungen basieren ebenfalls auf dem multiattributiven Präferenzmodell, bei dem

---

214 Zur Definition von Preisschwellen und damit verbundenen Effekten vgl. Gedenk/Sattler (1999).

215 Vgl. Metegrano (1994), S. 3-6.

216 Vgl. Srinivasan (1988), S. 296.

die erhobenen Präferenzen für die Eigenschaftsausprägungen mit den Wichtigkeiten der Eigenschaften multipliziert werden. Als Ergebnis der Self-Explicated-Phase liefert die ACA Teilnutzenwerte für alle Ausprägungen, deren Spannbreite das Bedeutungsgewicht einer Eigenschaft widerspiegeln. Darüber hinaus sind die Teilnutzenwerte so normiert, daß die Summe über jede Eigenschaft 0 ergibt.[217] Die formale Darstellung des Modells zur Schätzung der Teilnutzenwerte entspricht dem allgemeinen multiattributivem Präferenzmodell in Kapitel 2.1.[218]

3.3.2 Conjoint-Phase

Die Paarvergleichsaufgabe stellt das eigentliche Kernstück der ACA dar.[219] Für jeden Paarvergleich wird erfragt, welches der beiden Produktkonzepte der Proband bevorzugt und wie stark dieses bevorzugt wird. Die Präferenz wird auf einer 9-Punkte Rating-Skala quantifiziert und die Stimuli von 2 bis max. 5 Eigenschaften beschrieben.[220] Während eine größere Anzahl an Eigenschaften einen höheren Realitätsbezug der Produktbeschreibung bewirkt, steigt die Zahl zu verarbeitender Informationen. Aufgrund der damit verbundenen höheren Aufgabenkomplexität wird empfohlen, 2 oder max. 3 Eigenschaften zur Stimulibeschreibung heranzuziehen.[221] Dabei besteht die Möglichkeit die Zahl der Eigenschaften im Interviewablauf zu steigern und zunächst 2 und nach mehreren beantworteten Paarvergleichen, die folgenden Stimuli anhand von 3 Eigenschaften zu beschreiben.

Neben dieser geringeren Anzahl simultan zu beurteilender Stimuli unterscheidet sich die ACA sowohl in der Konzeption der Paarvergleiche (und damit hinsichtlich der Designkonstruktion) als auch in der Schätzung der Teilnutzenwerte gegenüber den beiden neuen computergestützten Verfahren. Die Gestaltung der Paarvergleiche wird während des gesamten Interviews von den Antworten des Probanden beeinflußt. Ausgehend von den Teilnutzenwerten der Self-Explicated-Phase werden die Paare so konstruiert, daß deren bis dahin ermittelte Gesamtnut-

---

217 Zur Berechnung der Teilnutzenwerte auf Basis der Self-Explicated-Phase vgl. Metegrano (1994), S. G-1 f.

218 Sofern die Präferenzen hinsichtlich der Eigenschaftsausprägungen mit einer Rangfolge erfaßt wurden, müssen diese zunächst in die umgekehrte Reihenfolge konvertiert werden.

219 Vgl. auch Green/Krieger/Agarwal (1991), S. 216 und Gutsche (1995), S. 96.

220 Vgl. Metegrano (1994), S. 3-7 f.

221 Vgl. Johnson (1987), S. 261.

zen möglichst ähnlich sind. Dieses Vorgehen dient wie die Pareto-Optimalität der Effizienzsteigerung der Datenerhebung. Je ähnlicher die Gesamtnutzen zweier Produktkonzepte sind, desto geringer ist die Wahrscheinlichkeit, daß eine Alternative eine andere eindeutig dominiert. Umgekehrt formuliert deuten große Unterschiede zwischen den Gesamtnutzen darauf hin, daß eine der beiden Alternativen eindeutig präferiert wird. Dieser Effekt tritt insbesondere dann auf, wenn die Alternativen überwiegend durch quantitative Eigenschaften beschrieben sind, bei denen die Präferenzen hinsichtlich der Ausprägungen bereits a priori definiert werden können. Nach jeder Bewertung eines Paarvergleichs werden die ermittelten Gesamtnutzen aktualisiert und damit die Wahl des nächsten Paarvergleichs determiniert. Gleichzeitig nimmt die Bedeutung der Ausgangsschätzungen auf Basis der Self-Explicated-Phase mit jedem zusätzlichen Paarvergleich ab. Das Zielkriterium, Paarvergleiche mit möglichst gleichen Gesamtnutzen zu konstruieren, wird unter der Nebenbedingung erfüllt, daß das Gesamtdesign näherungsweise orthogonal ist.

Verbunden mit dem adaptiven Charakter der Paarvergleiche unterscheidet sich die ACA auch wesentlich hinsichtlich der Schätzung der Teilnutzenwerte der dekompositionellen Phase. Sowohl bei der CCC als auch der CHIC basieren die Schätzergebnisse der Conjoint-Phase auf einer einmaligen Schätzung unter simultaner Auswertung aller erhobenen Informationen. Bei der ACA dagegen findet nach jedem Paarvergleich eine Aktualisierung der Schätzwerte statt. Als Ausgangspunkt dienen die auf Basis der Self-Explicated-Phase geschätzten Teilnutzenwerten, die mittels OLS-Schätzung mit den Schätzungen der Paarvergleiche kombiniert werden.[222] Diese werden mit jedem Paarvergleich aktualisiert und verfeinert, wobei das Gewicht der Self-Explicated-Teilnutzenwerte mit jedem zusätzlichen Paarvergleich geringer wird. Dieser Prozeß wird solange wiederholt bis ein entsprechend (vom Anwender) vorgegebenes Maß an Stabilität der Schätzung erfüllt ist oder die vorher festgelegte Zahl an Paarvergleichen beantwortet wurde. Durch die ständige Anpassung der Konstruktion der Paarvergleiche und die schrittweise Aktualisierung der Schätzergebnisse ist der Grad der Anpassung im Vergleich zu den beiden neuen Verfahren am höchsten. Während bei der CCC und CHIC lediglich eine Anpassung zwischen Self-Explicated-Phase und Con-

---

[222] Zum Schätzmodell für die Paarvergleichsaufgabe vgl. auch Green/Krieger/Agarwal (1991), S. 217.

joint-Phase stattfindet, wird diese bei der ACA auch innerhalb der Paarvergleichsaufgabe vorgenommen. Trotzdem ist es gerade die Paarvergleichsaufgabe, die aufgrund ihrer Ineffizienz in der Datenerhebung kritisiert wird.[223] Darüber hinaus bewerten Green/Krieger/Agarwal (1991) die Paarvergleichsaufgabe der ACA als sehr unrealistisch und durch die konstruierte Indifferenz zwischen den Konzepten für die Probanden sehr schwer zu bewältigen.[224]

Insgesamt kann an dieser Stelle festgehalten werden, daß sich die Conjoint-Phase der ACA in drei wesentlichen Merkmalen von den beiden neuen Verfahren CCC und CHIC unterscheidet. Zunächst sind jeweils nur 2 Stimuli in der dekompositionellen Phase simultan zu beurteilen. Des weiteren wird nicht nur zwischen Self-Explicated-Phase und Conjoint-Phase eine Anpassung vorgenommen, sondern bei der ACA ist auch die Paarvergleichsaufgabe in sich adaptiv gestaltet. Damit verbunden findet statt einer einmaligen Schätzung der Teilnutzenwerte eine kontinuierliche Anpassung der Schätzwerte statt, so daß nach dem letzten Paarvergleich Teilnutzenwerte vorliegen, die im Verlauf der dekompositionellen Phase entstanden bzw. ständig aktualisiert worden sind.

3.3.3 Kombinationsphase

Entsprechend der CCC und der CHIC liefert die ACA zwei getrennte Schätzungen für die beiden beschriebenen Interviewphasen. Zum einen die Schätzergebnisse auf Basis der Self-Explicated-Phase und zum anderen die im Verlauf der Paarvergleichsaufgabe aktualisierten Teilnutzenwerte. Die Kombinationsphase dient dazu, eine individuelle Gewichtung für diese beiden Ergebnisse zu bestimmen. Grundlage hierfür ist wiederum ein Validation-Set, mit dem das ACA-Interview abgeschlossen wird. Dieses Validation-Set besteht aus mindestens 4 bis max. 8 Produktkonzepten, wobei die Anzahl an Eigenschaften ebenfalls die Obergrenze von 8 nicht überschreiten darf. Jedes Produktkonzept wird für sich anhand einer Kaufwahrscheinlichkeit von 0% bis 100% bewertet, wobei zunächst das für die Testperson unattraktivste und anschließend das gemäß ihrer bisherigen Antworten attraktivste Produktkonzept zu bewerten ist. Alle weiteren Stimuli

---

223 Vgl. Green/Srinivasan (1990), S. 11; Louviere/Gaeth (1988), S. 63. Reiners (1996), S. 93 und Srinivasan/Park (1997), S. 286.

224 Vgl. Green/Krieger/Agarwal (1991), S. 220.

des Validation-Sets sind dann innerhalb dieser Bandbreiten zu bewerten, konsistentes Antwortverhalten vorausgesetzt.[225]

Das Zielkriterium für die Kombination der beiden Teilnutzenwerte ist, entsprechend dem Vorgehen bei der CCC und der CHIC, die Antworten hinsichtlich des Validation-Sets bestmöglich prognostizieren zu können. Die optimale Gewichtung der Teilnutzenwerte aus der Self-Explicated-Phase und der Paarvergleichsaufgabe wird bei der ACA mittels individueller OLS-Schätzung ermittelt. Als abhängige Variable geht dabei der Logit der Kaufwahrscheinlichkeit für die Produktbeschreibungen des Validation-Sets in das Modell ein[226]:

$$\text{(3-5)} \quad \ln\left[KW_i / \left(1 - KW_i\right)\right] \approx a_i + b_i \sum_{j \in J} P_{ijk_j}^{Pairs} + c_i \sum_{j \in J} P_{ijk_j}^{SE} \qquad (i \in I;\ k \in K_j)$$

wobei:

KW_i: Prognostizierte Kaufwahrscheinlichkeit für ein Produkt im Validation-Set durch den i-ten Probanden,

a_i: Konstantes Glied der Schätzgleichung für den i-ten Probanden,

b_i: zu schätzender Regressionskoeffizient für die Gewichtung der Teilnutzenwerte aus der Paarvergleichsaufgabe für den i-ten Probanden,

c_i: zu schätzender Regressionskoeffizient für die Gewichtung der Teilnutzenwerte aus der Self-Explicated-Phase für den i-ten Probanden,[227]

$P_{ijk_j}^{Pairs}$: Aktualisierter Teilnutzen nach dem letzten Paarvergleich der k-ten Ausprägung bei der j-ten Eigenschaft für den i-ten Probanden,

$P_{ijk_j}^{SE}$: Self-Explicated-Teilnutzen der k-ten Ausprägung bei der j-ten Eigenschaft für den i-ten Probanden.

---

225 Zur Gestaltung des Kalibrierungssets vgl. auch Reiners (1996), S. 120.

226 Zur Modellbeschreibung der Kombinationsphase vgl. Metegrano (1994), S. H-2.

227 Alternativ bietet die ACA 4.0 dem Anwender die Wahl, daß die Teilnutzenwerte aus den Paarvergleichen und der Self-Explicated-Phase gleichgewichtet aufsummiert werden. In diesem vereinfachten Fall wird nur ein Regressionskoeffizient für beide Teilnutzenwerte geschätzt. Formal gilt in diesem Fall $b_i = c_i$.

Die Regressionskoeffizienten b_i und c_i entsprechen inhaltlich den Gewichtungsfaktoren w_i^* und $(1-w_i^*)$ bei der CCC und CHIC. Der Unterschied besteht in der jeweiligen Ermittlung dieser Gewichtungsfaktoren. Während die Werte bei der ACA auf Basis einer individuellen OLS-Schätzung bestimmt werden, basiert die optimale Gewichtung bei der CCC und CHIC auf einer optimalen Kreuzvalidierung zwischen beobachteten und prognostizierten Präferenzen hinsichtlich des Validation-Sets, gemessen durch entsprechende Korrelationskoeffizienten.[228]

Damit unterscheidet sich auch die Kombinationsphase der ACA in zwei wesentlichen Merkmalen von den beiden neuen Methoden. Aus Sicht der Datenerhebung wird deutlich, daß die Beurteilungsaufgabe für den Probanden in der Kombinationsphase eine andere ist als in der Conjoint-Aufgabe. Während dort Paarvergleiche, beschrieben durch 2 bis 3 Eigenschaften, vorgelegt werden, umfaßt das Validation-Set nahezu Vollprofile, die alle separat zu bewerten sind. Im Gegensatz dazu ist die Beurteilungsaufgabe im Validation-Set bei der CCC und der CHIC jeweils identisch mit der Abfrage in der eigentlichen Conjoint-Aufgabe. Der zweite Unterschied betrifft die optimale Kombination der Teilnutzenwerte aus beiden Datenerhebungsphasen. Während bei der CCC und CHIC die optimalen Gewichtungsfaktoren durch Maximierung der Korrelationskoeffizienten zwischen beobachteten und prognostizierten Präferenzen ermittelt werden, schätzt die ACA diese Werte mittels individueller OLS-Schätzung.

Die Struktur insbesondere der Datenerhebung der ACA wird im folgenden Ablaufschema (vgl. Abbildung 3-5) zusammenfassend dargestellt.

---

[228] Zur Schätzung der Teilnutzenwerte bei der ACA vgl. Sawtooth Software (1996), S. 16 ff.

Abbildung 3-5: Interviewablauf der ACA

Phase	Befragungsschritt	Vorgehensweise bei der ACA
Self-Explicated-Phase	1. Schritt: Eliminierung inakzeptabler Eigenschaftsausprägungen	• Direkte Frage nach inakzeptablen Eigenschaftsausprägungen (optionaler Befragungsschritt)
	2. Schritt: Erfassung von Präferenzen gegenüber Eigenschaftsausprägungen	• Angabe einer Präferenzrangfolge für die Ausprägungen einer Eigenschaft oder Beurteilung der Ausprägungen anhand einer Rating-Skala (2 bis 9 Kategorien)
	3. Schritt: Erfassung der Wichtigkeiten	• Frage nach der Wichtigkeit separat bei allen Eigenschaften anhand einer Rating-Skala (2 bis 9 Kategorien)
	➔ Schätzung der Teilnutzenwerte führt zur **SE-ACA**	
Conjoint-Phase	Erfassung ganzheitlicher Präferenzen	• Frage nach der Präferenz und deren Stärke zwischen Produktpaaren (jeweils beschrieben anhand von 2 bis 5 Eigenschaften) auf einer Rating-Skala (2 bis 9 Kategorien)
Kombinationsphase	Kombination der Informationen aus den beiden vorangegangenen Befragungsteilen	• Frage nach der Kaufwahrscheinlichkeit (0 bis 100%) bei 4 bis 8 Produkten, die anhand von 2 bis 8 Eigenschaften beschrieben sind
	➔ Kombination der Teilnutzenwerte gemäß individueller OLS-Schätzung führt zur **ACA**	

Quelle: Eigene Erstellung

Nachdem die 3 computergestützten hybriden Verfahren CCC, CHIC und ACA separat dargestellt wurden, faßt die folgende Abbildung 3-6 alle wesentlichen Merkmale und Unterschiede der 3 Methoden noch einmal zusammen. Schwerpunkt der Gegenüberstellung sind die Unterschiede in der Datenerhebung als auch in der Schätzung der Teilnutzenwerte der Self-Explicated- und der Conjoint-Phase sowie deren Kombination. Die Übersicht ist entsprechend der 3 Phasen (Self-Explicated-, Conjoint- und Kombinationsphase) unterteilt.

Abbildung 3-6: Merkmale der untersuchten computergestützten hybriden Verfahren im Vergleich (Teil 1)

Merkmal	**CCC**	**CHIC**	**ACA**
Self-Explicated-Phase			
Eliminierung inakzeptabler Eigenschaften	•noch nicht implementiert		•optionaler Befragungsschritt
Bewertung der Eigenschaftsausprägungen	•Abfrage auf einer 11-Punkte-Rating-Skala •A priori - Definition der Extremwerte •Sprünge in der Teilnutzenfunktion können berücksichtigt werden		•Bewertung alternativ mittels Ranking oder Rating-Skala (2 bis 9 Kategorien) •A priori - Definition nur für alle Ausprägungen möglich
Erfassung der Wichtigkeiten	•Bewertung aller Eigenschaften im Vergleich zueinander unter Verwendung der wichtigsten Eigenschaft als Ankerpunkt		•Alle Eigenschaften werden separat bewertet
Conjoint-Phase			
Anzahl simultan zu beurteilender Stimuli	•alle zu beurteilenden Stimuli: **8 bis 16**	•Je nach Größe der Choice-Sets: **4 bis 5**	•Paarvergleiche: **2**
Anzahl der Eigenschaften zur Stimulibeschreibung	•**3 bis 6** Kerneigenschaften	•**3 bis 6** Kerneigenschaften	•**2 bis 5** Eigenschaften
Beurteilungsaufgabe	•**Rangreihung** und/oder **Rating**-Skala	•Wiederholte **Auswahlentscheidungen** •Quantifizierung der Präferenzunterschiede •Transformation der Präferenzunterschiede in eine monetäre Größe	•Gewichtete (Graded) **Paarvergleiche**
Design zur Stimulikonstruktion	•Orthogonales Design	•Orientierung an Pareto-Optimalität	•gebildet werden möglichst nutzengleiche Paare •Orthogonalität als Nebenbedingung
Ausmaß der Anpassung (Adaptivität)	•einmalig zwischen Self-Explicated- und Conjoint-Phase		•zwischen Self-Explicated-Phase und Conjoint-Phase und innerhalb der Paarvergleichsaufgabe
Schätzung der Teilnutzenwerte	•einmalige Schätzung mittels LINMAP auf Basis der Rangdaten bzw. mittels OLS-Schätzung auf Basis der Rating-Daten	•einmalige Schätzung mittels OLS auf Basis der metrisch skalierten Präferenzdaten	•schrittweise Aktualisierung der Ausgangsschätzung nach jedem Paarvergleich

Abbildung 3-6: Merkmale der untersuchten computergestützten hybriden Verfahren im Vergleich (Teil 2)

Merkmal	CCC	CHIC	ACA
Kombinationsphase			
Größe des Validation-Sets	•3 bis 12 Produkte	•je nach Größe der Choice-Sets 4 bis 6	•4 bis 8
Beurteilung des Validation-Sets	•identisch mit Conjoint-Phase		•Abfrage der Kaufwahrscheinlichkeit von 0% bis 100% für ganzheitlich beschriebene Produkte
Ermittlung der optimalen Gewichtungsfaktoren	•Gemessen durch maximale Korrelation zwischen beobachteter und prognostizierter Präferenzen für das Validation-Set		•Individuelle OLS-Schätzung mit Gewichtungsfaktoren als Regressionskoeffizienten

Quelle: Eigene Erstellung

Faßt man die Hauptunterschiede zwischen den Verfahren zusammen, werden im wesentlichen folgende zwei Punkte deutlich. Mit der CCC und der CHIC werden zwei neue computergestützte hybride Verfahren vorgestellt, bei denen jeweils mehr als 2 Stimuli in der Conjoint-Phase des Interviews simultan zu bewerten sind. Damit wird der Bereich computergestützter hybrider Verfahren um neue Methoden erweitert, bei denen alternative dekompositionelle Ansätze berücksichtigt werden. Eine vergleichende Validitätsprüfung bietet dann die Möglichkeit zu testen, ob das relativ schlechte Abschneiden der ACA in bisherigen empirischen Untersuchungen u.a. in der häufig kritisierten Konzeption der Paarvergleichsaufgabe begründet liegt.

Der zweite wesentliche Unterschied zwischen den Verfahren betrifft das Ausmaß der Individualisierung der einzelnen Verfahren. Auch wenn alle 3 vorgestellten Verfahren einen adaptiven Charakter aufweisen, variiert dabei das Ausmaß, in dem nachfolgende Befragungsschritte an bereits gewonnene Informationen angepaßt werden. So findet sowohl bei der CCC als auch der CHIC eine einmalige Anpassung zwischen der Self-Explicated- und der Conjoint-Phase statt. Bei der ACA dagegen werden darüber hinaus auch innerhalb der Conjoint-Phase die einzelnen Paarvergleiche in Abhängigkeit bereits erhobener Präferenzurteile konzipiert. Gleichzeitig findet hierbei eine kontinuierliche Schätzung der Teilnutzenwerte innerhalb des Conjoint-Interviews statt. Die Conjoint-Teilnutzenwerte der

CCC und der CHIC werden im Gegensatz hierzu durch eine einmalige Schätzung bestimmt, bei der alle im Rahmen der Conjoint-Phase erhobenen Präferenzdaten gemeinsam ausgewertet werden. Die Verfahren weichen somit neben der unterschiedlichen Gestaltung der Conjoint-Phase dahingehend voneinander ab, inwieweit diese individualisiert und an vorangegangene Antworten angepaßt wird. Ob und in welcher Form sich diese Unterschiede auf die Validität der Ergebnisse auswirken, muß empirisch überprüft werden. Im Rahmen der vorliegenden Arbeit wurde aus diesem Grund eine vergleichende Studie durchgeführt, deren Untersuchungsdesign im nächsten Kapitel beschrieben wird.

4 Untersuchungsdesign

Als zweites Ziel der vorliegenden Arbeit sollen die neu entwickelten Methoden einer empirischen Validitätsprüfung unterzogen werden. Zunächst wird in diesem Kapitel das Untersuchungsdesign der hierzu durchgeführten empirischen Studie beschrieben. Neben der Darstellung des Untersuchungsgegenstands wird die Zusammensetzung der Gesamtstichprobe dokumentiert und die Teilstichproben auf strukturelle Vergleichbarkeit getestet. Im dritten Abschnitt des Kapitels werden der Interviewablauf der Befragung und die Aufbereitung der Daten kurz skizziert.

4.1 Untersuchungsobjekt

Gegenstand der empirischen Untersuchung sollte ein Produkt sein, das von einer großen Anzahl potentieller Nachfrager beurteilt werden kann. Vor diesem Hintergrund wurden **Kühlschränke** als Untersuchungsgegenstand gewählt. Das Produkt „Kühlschrank" stellt einen sehr weit verbreiteten Gebrauchsgegenstand dar, dessen Eigenschaften größtenteils auch von bisherigen Nichtkäufern beurteilt werden können. Aufgrund hoher Bekanntheit und relativ hohem Involvement beim Kauf des Produkts kann davon ausgegangen werden, daß die Bewertungen der verschiedenen Eigenschaften bzw. deren Ausprägungen nicht willkürlich getroffen werden.[229] Gleichzeitig handelt es sich bei Kühlschränken um ein Produkt, bei dem eine größere Anzahl an Eigenschaften für die Kaufentscheidung relevant ist. Auch dies war entscheidend für die Produktwahl, da die getesteten Methoden gerade für solche Fragestellungen Vorteile aufweisen. Darüber hinaus wurden bereits in anderen Studien „Kühlschränke" als Beispiel herangezogen. So unter anderem in einem sehr umfassenden Vergleich alternativer Präferenzmeßmethoden von Huber et al. (1993).[230] Die Wahl des gleichen Produkts erhöht somit die prinzipielle Vergleichbarkeit der Untersuchungsergebnisse.

Aufgrund zahlreicher unterschiedlicher Modelle und Typen innerhalb der Produktklasse „Kühlschrank", wurde für die empirische Untersuchung eine Einschränkung auf sogenannte Tischmodelle (Maße: 80cm/60cm/60cm) mit Gefrierfach vorgenommen. Dies war notwendig, um bei systematischer Variation der

---

229 Vgl. hierzu auch Hahn (1997), S. 161.

230 Vgl. Huber et al. (1993), S. 107.

Eigenschaftsausprägungen insgesamt realistische und vergleichbare Eigenschaftskombinationen zu erhalten. Auf Basis von Produktbeschreibungen der Hersteller sowie in Anlehnung an bisher veröffentlichte Studien mit Kühlschränken als Untersuchungsgegenstand wurden zunächst 13 Eigenschaften mit 2 bis 4 Ausprägungen betrachtet. Im Rahmen einer schriftlichen Befragung von 32 Auskunftspersonen sind die Wichtigkeiten dieser 13 Eigenschaften anhand einer 7-Punkte-Rating-Skala (1 = "vollkommen unwichtig"; 7 = "sehr wichtig") erfragt worden. Nach diesem Pretest konnte eine Reduzierung der Merkmale vorgenommen werden. Für die Auswahl einer Eigenschaft war entscheidend, daß sich über alle Auskunftspersonen ein Mittelwert der Wichtigkeitsurteile > 4 ergab.[231] Das Ergebnis des Pretests zeigte, daß genau 8 der 13 Eigenschaften diese Bedingung erfüllten. Die Zusammenstellung der Eigenschaftsausprägungen erfolgte unter dem Gesichtspunkt, für das betrachtete Preissegment möglichst realistische Kombinationen aller anderen Eigenschaften vornehmen und die real am Markt existierenden Modelle abbilden zu können. Die untersuchten 8 Eigenschaften und ihre Ausprägungen sind in Tabelle 4-1 wiedergegeben.[232]

---

231 Der Wert 4 bildet die Mitte der angewendeten 7-Punkte-Rating-Skala mit den Extrempolen 1 = „vollkommen unwichtig" und 7 = „sehr wichtig". Ein Mittelwert >4 impliziert demnach, daß für die Mehrheit der Auskunftspersonen die entsprechende Eigenschaft eher wichtig als unwichtig ist.

232 Die fünf Eigenschaften, die sich im Rahmen des Pretests als weniger wichtig erwiesen, waren: „Temperaturanzeige", „Lagerzeit bei Stromausfall", „Super-Frost-Funktion", „Material der Abstellflächen" und „Automatische Türschließung".

Tabelle 4-1: Untersuchte Eigenschaften und deren Ausprägungen

Eigenschaft	Ausprägungen
Marke	• Liebherr • Siemens • Electrolux • Quelle Privileg
Preis	• 449,-- DM • 549,-- DM • 649,-- DM
Energieverbrauch pro Jahr (Energie)	• 150 kWh pro Jahr • 200 kWh pro Jahr • 250 kWh pro Jahr
Garantiedauer (Garantie)	• 12 Monate • 18 Monate • 24 Monate
Nutzinhalt (Inhalt)	• 115 Liter • 130 Liter • 145 Liter
Innenausstattung (Innen)	• Innenaufteilung vorgegeben • Kühlteil variabel • Kühlteil und Tür variabel
Abtaumöglichkeit (Abtau)	• von Hand • Abtau-Automatik im Kühlteil
Sterne des Gefrierfachs (Gefrier)	• *-Fach (-6° C) • ***-Fach (-18°C)

Quelle: Eigene Erstellung[233]

Die nicht berücksichtigten Eigenschaften wurden für alle zu bewertenden Produktbeschreibungen konstant gehalten und die Testpersonen in den einzelnen Phasen des Interviews auf diese Tatsache aufmerksam gemacht.

4.2 Stichprobe

Die empirische Überprüfung der Methoden basiert auf einer jeweils unabhängigen Stichprobe potentieller Nachfrager für Kühlschränke für die ACA, CCC und CHIC (between-subject-design). Durch Auswahl einer separaten Stichprobe für die einzelnen hybriden Methoden wird vermieden, daß die Ergebnisse durch Rei-

[233] Es sei darauf hingewiesen, daß alle Tabellen und Abbildungen, die sich auf die Ergebnisse der vorliegenden Untersuchung beziehen, von der Autorin erstellt wurden. Auf eine wiederholte Kennzeichnung wird im folgenden verzichtet.

henfolge- und/oder Lerneffekte bei den Probanden beeinflußt werden.[234] Außerdem kann so eine Überforderung der Auskunftsbereitschaft der Probanden vermieden werden.[235] Die Auswahl der Personen für die Befragung erfolgte nach dem Quotenverfahren, wobei „Alter" und „Geschlecht" als Quotenmerkmale dienten. Die Ausprägungen der Quotenmerkmale wurden entsprechend des Bundesdurchschnitts der Gesamtbevölkerung festgelegt. Zudem wurden nur diejenigen Personen befragt, die sich prinzipiell vorstellen konnten, in den nächsten 12 Monaten einen Tischkühlschrank zu kaufen.[236] Mit dieser zusätzlichen Eingrenzung des Personenkreises soll sichergestellt werden, daß die Probanden zur Zielgruppe für diese Größe und Art von Kühlschränken zählen. Einschränkend muß an dieser Stelle angemerkt werden, daß die gewählte Stichprobe nur näherungsweise als repräsentativ für die betrachtete Zielgruppe angesehen werden darf. Zum einen wurden lediglich zwei Quotenmerkmale zur Auswahl der Probanden herangezogen und zum anderen kann nicht uneingeschränkt davon ausgegangen werden, daß deren Ausprägungen gemäß dem Bundesdurchschnitt den potentiellen Nachfragern für Tischkühlschränke entsprechen.

Die Gesamtstichprobe umfaßte insgesamt 748 Personen und stellt damit eine der bisher umfassendsten wissenschaftlichen Studien hinsichtlich eines Methodenvergleichs dar.[237] Alle Interviews wurden im Juni 1998 von dem Marktforschungsinstitut IPSOS Deutschland durchgeführt.[238] Es handelte sich dabei um Studio-Interviews in den Städten Hamburg, Frankfurt/Main und München. In jeder Stadt wurden alle drei Methoden CCC, CHIC und ACA gleich häufig eingesetzt. Von der Gesamtzahl durchgeführter Interviews konnten letztendlich nur die in die Analyse einbezogen werden, die vollständig beantwortet bzw. auszuwerten waren. Nicht vollständig beantwortet wurden 23 (ACA:1; CCC: 8; CHIC: 14) Interviews, so daß der empirische Validitätsvergleich auf insgesamt 725 (ACA: 249: CCC: 242; CHIC: 234) ausgewerteten Interviews basiert. Einen Über-

---

234 Vgl. Agarwal/Green (1991), S. 145; Huber et al. (1993), S. 110; Johnson (1991), S. 223.

235 Vgl. Sattler (1994), S. 34.

236 Um zu verdeutlichen, um welche Art von Kühlschränken es sich bei der Kategorie „Tischkühlschränke" handelt, wurden den Personen Fotos verschiedener Modelle vorgelegt.

237 Vgl. hierzu die Stichproben bisheriger Vergleichsstudien in Kapitel 2.

238 An dieser Stelle sei der IPSOS Deutschland, namentlich Herrn Harald Hasselmann und Frau Claudia Keck für die Durchführung der Datenerhebung und die außerordentlich gute Zusammenarbeit gedankt.

blick über den Umfang und die demografische Struktur der drei Teilstichproben gibt nachstehende Tabelle. Sofern einzelne Probanden keine Angabe zu den erhobenen demografischen Merkmalen gemacht haben, werden diese als „missing values“ aufgeführt.

Tabelle 4-2: Umfang und demografische Struktur der Teilstichproben

Methode	**ACA**		**CCC**		**CHIC**	
	absolut	in %	absolut	in %	absolut	in %
Demografische Merkmale	n=**249**	n=100%	n=**242**	n=100%	n=**234**	n=100%
Geschlecht:						
männlich:	122	49	114	47	118	50
weiblich:	127	51	127	53	116	50
missing values	-	-	1	0	-	-
Alter:						
18-29 Jahre	54	22	54	22	54	23
30-39 Jahre	65	26	66	27	73	31
40-49 Jahre	42	17	38	16	24	10
50-70 Jahre	88	35	84	35	83	35
Bildung:						
Hauptschule	9	4	4	2	4	2
Hauptschule mit Berufsausbildung	37	15	32	13	46	20
mittlere Reife	92	37	88	36	78	33
Abitur	73	29	63	26	63	27
Universität	38	15	53	22	40	17
missing values:	-	-	2	1	3	1
Haushaltsnetto-einkommen:						
bis 1.999,- DM	81	33	37	15	39	17
2.000-2.999,- DM	48	19	47	19	48	21
3.000-3.999,- DM	13	5	37	15	36	15
4.000-4.999,- DM	39	16	30	12	32	14
≥ 5.000,- DM	41	16	53	22	53	23
missing values	27	11	34	14	25	11
Haushaltsgröße:						
1 Person	99	40	90	37	86	37
2 Personen	77	31	82	34	89	38
3 Personen	33	13	31	13	24	10
4 Personen	27	11	26	11	29	12
5 und mehr Personen	13	5	13	5	6	3

Datenquelle: IPSOS Deutschland (1998)

Voraussetzung für einen Vergleich der Ergebnisse zwischen den einzelnen Methoden, die auf unterschiedlichen Teilstichproben basieren, ist die Vergleichbar-

keit dieser Stichproben. Nur wenn diese gegeben ist, kann ausgeschlossen werden, daß differierende Ergebnisse lediglich auf unterschiedliche Charakteristika der Stichproben zurückgeführt werden können bzw. müssen. Zur Überprüfung der Strukturgleichheit der drei Sample wurde jeweils paarweise zwischen den Methoden ein χ^2-Homogenitätstest[239] im Hinblick auf die sozio-demografischen Merkmale Alter, Geschlecht, Bildungsabschluß, Haushaltsnettoeinkommen und Haushaltsgröße durchgeführt. Da die beiden Merkmale Alter und Geschlecht als Quotenmerkmale herangezogen wurden, war die Vergleichbarkeit der Stichproben hinsichtlich dieser beiden Merkmale zu erwarten.[240] Die Informationen zu den drei anderen Merkmalen wurden erst am Ende des Interviews erhoben und die Teilstichproben nachträglich auf Vergleichbarkeit hinsichtlich dieser Merkmale überprüft. Die Ergebnisse des Homogenitätstests sind in Tabelle 4-3 zusammengefaßt. In den Spalten stehen die Ergebnisse der verschiedenen Paarvergleiche zwischen den jeweiligen Teilstichproben für das ACA, CCC und CHIC-Verfahren.

Tabelle 4-3: Testergebnisse der Überprüfung der Teilstichproben auf Strukturgleichheit #

Merkmal	**ACA vs. CCC**	**ACA vs. CHIC**	**CCC vs. CHIC**
Geschlecht	n.s.	n.s.	n.s.
Alter	n.s.	n.s.	n.s.
Bildung	n.s.	n.s.	n.s.
Einkommen	29,94 ***	27,29 ***	n.s.
Haushaltsgröße	n.s.	n.s.	n.s.

Prüfung anhand paarweiser χ^2-Homogenitätstests. Bei signifikanten ($p<0,10$) Unterschieden werden sowohl der χ^2-Wert als auch das Signifikanzniveau ausgewiesen.
n.s.: nicht signifikant ($p>0,10$); * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$

Als Ergebnis der durchgeführten χ^2-Tests zeigt sich, daß auch bei einer Irrtumswahrscheinlichkeit von $p>0,10$ die Nullhypothese der Strukturgleichheit in 13 der 15 durchgeführten Vergleiche nicht abgelehnt werden kann. Einzige Ausnahme bildet das Merkmal Einkommen. Hinsichtlich dieses Merkmals weicht die Stich-

239 Zum χ^2-Homogenitätstest bei kategorialen Daten vgl. Falk/Becker/Marohn (1995), S. 137 ff.

240 Die aufgetretenen (geringen) Abweichungen zwischen den drei Teilstichproben hinsichtlich dieser beiden Merkmale ergaben sich dadurch, daß unvollständige Interviews aus der Analyse ausgeschlossen werden mußten.

probe für die ACA signifikant von den beiden anderen Teilstichproben ab. Diese Abweichung ist hauptsächlich auf die vergleichsweise große Anzahl befragter Personen der untersten Einkommenskategorie im Sample der ACA zurückzuführen (vgl. hierzu Tabelle 4-2). Eine genauere Analyse dieser demografischen Variable ergab jedoch, daß sie keinen signifikanten Einfluß auf die Validität der Ergebnisse ausübt.[241] Insgesamt kann somit strukturelle Vergleichbarkeit zwischen den unabhängigen Teilstichproben vermutet werden.

4.3 Interviewablauf

Während in Kapitel 3 die untersuchten Verfahren konzeptionell dargestellt wurden, ist das Ziel der folgenden Ausführungen, den Interviewablauf und die Datenauswertung der vorliegenden empirischen Anwendung zu dokumentieren. Dabei werden die einzelnen Verfahrensschritte nicht noch einmal ausführlich erläutert (vgl. hierzu Kapitel 3), sondern die Besonderheiten bei der konkreten Umsetzung der einzelnen Befragungsschritte beschrieben. Die Erläuterungen orientieren sich an der bereits in Kapitel 3 vorgenommenen Strukturierung des Interviews in Self-Explicated-, Conjoint- und Kombinationsphase.

4.3.1 Self-Explicated-Phase

In der Self-Explicated-Phase aller 3 Verfahren beurteilten die Probanden zunächst die 8 Eigenschaften aus Tabelle 4-1 entsprechend der in Kapitel 3 erläuterten Vorgehensweise.[242] Bei den quantitativen Eigenschaften Preis, Energieverbrauch, Garantieleistung und Nutzinhalt wurde für die Bewertung der Ausprägungen bereits eine a priori Ordnung vorgenommen. Beispielsweise gilt, daß bei angenommener Konstanz aller anderen Eigenschaften der niedrigste (höchste) Energieverbrauch als beste (schlechteste) Ausprägung für alle Probanden festgelegt wurde. Die Integration solcher a priori Informationen erhöht nicht nur die Effizienz der Datenerhebung, sondern hilft, Schätzfehler zu vermeiden und folglich die Validität der Ergebnisse zu verbessern.[243] Entsprechend entfielen bei der CCC und CHIC die Abfrage nach jeweils bester und schlechtester Ausprägung

---

241 Vgl. Sattler/Hensel-Börner/Krüger (2000), S. 10. Des weiteren sei angemerkt, daß aufgrund einer relativ hohen Anzahl an missing values, die Aussagekraft dieser Variable für das Gesamtsample eingeschränkt ist.

242 Die Abfrage inakzeptabler Eigenschaftsausprägungen wurde in der vorliegenden Programmversion der CCC und der CHIC noch nicht implementiert. Aus Gründen der Vergleichbarkeit wurde auch bei der ACA auf diesen Befragungsschritt verzichtet.

243 Vgl. Srinivasan/Jain/Malhotra (1983), S. 434.

für diese Eigenschaften. Die Probanden hatten lediglich die mittleren Ausprägungen auf einer 11-stufigen Rating-Skala zu bewerten, an deren Endpole die beste und schlechteste Ausprägung als Ankerpunkte fungierten. Für die ACA bedeutet eine solche a priori Definition von Eigenschaftsausprägungen, daß die Bewertung für alle Ausprägungen dieser Eigenschaft entfällt. Die Ausprägungen der verbleibenden qualitativen Eigenschaften wurden bei der ACA anhand einer 9-Punkte-Skala beurteilt.[244] Im nächsten Befragungsschritt wird jeweils die gleiche 11- bzw. 9-Punkte-Skala verwendet, um die Wichtigkeiten der Eigenschaften zu erfassen. Hierfür wurden die Eigenschaften immer mit der individuell besten und schlechtesten Ausprägung auf dem Bildschirm präsentiert. Bei der CCC und CHIC bestand die Aufgabe der Probanden darin, alle 8 Eigenschaften simultan und unter Verwendung der individuell wichtigsten Eigenschaft als Ankerpunkt zu bewerten. Im Gegensatz dazu wurden die Eigenschaften bei der ACA jeweils einzeln präsentiert und damit separat voneinander beurteilt.[245] Unter Annahme eines additiven Nutzenmodells wurden für alle drei Verfahren in gleicher Form die Self-Explicated-Teilnutzenwerte ermittelt. Diese isolierte Betrachtung der ersten Phase ist Grundlage des Methodenvergleichs zwischen den hybriden Verfahren mit den einfacheren Self-Explicated-Methoden. Im folgenden werden die Ergebnisse der Self-Explicated-Verfahren mit **SE-CCC**, **SE-CHIC** und **SE-ACA** ausgewiesen.

4.3.2 Conjoint-Phase

CCC

Der dekompositionelle Teil der CCC umfaßte die Beurteilung von 9 verschiedenen Kühlschränken, die anhand der individuell wichtigsten 4 Eigenschaften beschrieben wurden. Während die Anzahl von 4 Kerneigenschaften entsprechend der Studie von Srinivasan/Park (1997) festgelegt wurde,[246] stellen 9 Stimuli einen Kompromiß zwischen ausreichender Anzahl an Beobachtungen und einer auf dem Bildschirm darstellbaren Obergrenze an Produktprofilen dar. Die Ausprägungen dieser Kerneigenschaften wurden systematisch zwischen den Produkten

---

[244] Für die ACA ist dies die größtmögliche Anzahl an Skalenpunkten für eine Rating-Skala (vgl. Metegrano (1994), S. 3-31). Diese wurde gewählt, um die bestmögliche Vergleichbarkeit zur 11-Punkte-Rating-Skala der SE-CCC und SE-CHIC zu erzielen.

[245] Zu den Problemen dieser separaten Abfrage der Wichtigkeiten vgl. Reiners (1996), S. 118.

[246] Vgl. Srinivasan/Park (1997), S. 288.

gemäß eines orthogonalen Untersuchungsdesigns variiert.[247] Die Beurteilungsaufgabe der Probanden bestand aus zwei Teilschritten. Zunächst waren die 9 auf dem Bildschirm dargestellten Kühlschränke in die Rangfolge zu bringen, in der die Probanden die Modelle am ehesten kaufen würden. Entsprechend der angegebenen Rangfolge sortierte das Programm die Kühlschränke anschließend auf dem Bildschirm. Im zweiten Teilschritt sollte für jeden Kühlschrank eine Kaufwahrscheinlichkeit von 0% bis 100% angegeben werden. Während Rangplätze nicht doppelt vergeben werden konnten, war die Angabe gleicher Kaufwahrscheinlichkeiten für verschiedene Produkte grundsätzlich möglich. Sichergestellt war aber, daß einem Kühlschrank mit einem höherem Rang keine höhere Kaufwahrscheinlichkeit zugeordnet werden konnte.[248] Die Teilnutzenwerte wurden auf Basis der Ranking-Daten mit ordinalem Skalenniveau mit LINMAP von Bretton-Clark und unter Verwendung der Kaufwahrscheinlichkeiten mit metrischem Skalenniveau mittels OLS-Schätzung geschätzt.[249] Um die Anzahl zu schätzender Parameter so gering wie möglich zu halten, wurden bei beiden Schätzverfahren für alle quantitativen Eigenschaften Vektor- und nur für die qualitativen Eigenschaften Teilnutzenmodelle unterstellt. Das Verfahren LINMAP ermöglicht hierbei, a priori Ordnungen für Eigenschaftsausprägungen bei der Schätzung zu berücksichtigen. Die bereits im Self-Explicated-Teil erhobenen Präferenzen hinsichtlich der Eigenschaftsausprägungen wurden daher bei der Schätzung der Teilnutzenwerte mit LINMAP als Nebenbedingung für das Optimierungsverfahren integriert.[250] Um die ermittelten Conjoint-Teilnutzenwerte für die Kerneigenschaften mit den Ergebnissen der Self-Explicated-Phase vergleichen zu können, war eine Umskalierung der Conjoint-Ergebnisse in der Weise notwendig, daß die Summe der Be-

---

247 Zu orthogonalen Designplänen vgl. Addelman (1962), S. 34 ff.

248 Auf dem Bildschirm erschien in diesem Fall eine entsprechende Fehlermeldung mit der Bitte an den Probanden, seine Eingabe zu korrigieren.

249 Das Verfahren LINMAP ermöglicht grundsätzlich zwar auch die Auswertung von Rating-Daten, verwendet für die Schätzung der Teilnutzenwerte jedoch lediglich die ordinalen Informationen hinsichtlich der erhobenen Präferenzen (vgl. Bretton-Clark (1989), S. 50). Die durch den zweiten Teilschritt zusätzlich erhobenen Informationen fänden daher keine Berücksichtigung in der Schätzung. Demzufolge würde eine Auswertung der Kaufwahrscheinlichkeitsdaten mit diesem Verfahren zu den gleichen Ergebnissen wie auf Basis der Rangdaten führen. Abweichungen konnten nur bei den Probanden auftreten, die die gleichen Kaufwahrscheinlichkeiten mehrfach vergeben haben.

250 Van der Lans/Heiser (1992) zeigen empirisch, daß die Integration solcher „constraints" bei der Schätzung der Conjoint-Teilnutzenwerte die Schätzergebnisse, gemessen im Rahmen einer Kreuzvalidierung, verbessert werden können. Die Verbesserungen fallen zwar klein aus, sind aber statistisch signifikant (vgl. van der Lans/Heiser (1992), S. 336).

deutungsgewichte für die vier Kerneigenschaften gleich der entsprechenden Summe der Self-Explicated-Ergebnisse entsprach.[251]

CHIC

Für die CHIC bestand die Aufgabe der Conjoint-Phase in der Bewertung verschiedener Choice-Sets. Die Anzahl insgesamt zu beurteilender Choice-Sets hing dabei von den individuellen Kerneigenschaften ab. Analog zur CCC wurden die Produktkonzepte anhand der 4 individuell wichtigsten Eigenschaften beschrieben. Zählten beide Eigenschaften mit nur 2 Ausprägungen (Abtaumöglichkeit und Sterne des Gefrierfachs) zu den Kerneigenschaften, waren in der Conjoint-Phase 2, andernfalls 3 Choice-Sets zu bewerten. Ein Choice-Set bestand aus jeweils 4 Produktalternativen und der Nicht-Kauf-Alternative. Wie in Kapitel 3.2 dargestellt, stand bei der Konstruktion dieser Choice-Sets das Prinzip der Pareto-Optimalität im Vordergrund.[252] Hierfür wählte das Programm aus dem jeweils vollständigen Design zunächst alle diejenigen Eigenschaftskombinationen aus, die sich in mindestens 2 Eigenschaften unterschieden. Damit reduzierte sich die Zahl zu bewertender Stimuli bereits erheblich, denn je nach individuellen Kerneigenschaften und dem daraus entstehenden vollständigen Design aus 81 (3*3*3*3), 54 (3*3*3*2) oder 36 (3*3*2*2) Eigenschaftskombinationen, erfüllten nur noch 22, 14 oder 10 Produktbeschreibungen diese Nebenbedingung.[253] Die anschließende Aufteilung der Stimuli auf die verschiedenen Choice-Sets erfolgte zufällig, um systematische Verzerrungen zu vermeiden. Folglich variiert die Zusammenstellung der Choice-Sets über die Probanden, wodurch das Ausmaß der Individualisierung des Interviewablaufs erhöht wird. Nachteilig wirkt sich diese zusätzliche Individualisierung allerdings auf den Aufwand der Datenauswertung aus, da noch keine Standardprogramme für die individuelle Auswertung bei den neuen Methoden vorliegen.

Die Gestaltung der Beurteilungsaufgabe entsprach der in Kapitel 3.2.2.2 erläuterten Vorgehensweise. Durch eine wiederholte Auswahlentscheidung wurden

---

251 Vgl. Srinivasan/Park (1997), S. 288.

252 Zu den Eigenschaften eines pareto-optimalen Designs vgl. Kapitel 3.2.2.1 dieser Arbeit.

253 Um die Größe der Choice-Sets konstant halten zu können, wurde bei einem 3*3*3*2-Design (3*3*2*2-Design) der Kerneigenschaften das bereits reduzierte Design von 14 (10) Stimuli um zwei weitere ergänzt, um 4 (3) gleiche große Choice-Sets bilden zu können. Um hierbei systematische Einflüsse zu vermeiden, wurden auch diese beiden Ergänzungskarten jeweils zufällig aus dem vollständigen Design ausgewählt.

die Alternativen eines Choice-Sets zunächst in eine Präferenzrangfolge gebracht. Die Präferenzunterschiede sollten anschließend anhand einer 100 Punkte-Skala bewertet werden, bei der die zuerst (100 Punkte) und zuletzt (0 Punkte) ausgewählten Alternativen die Endpole bildeten. Für den letzten Teilschritt der Beurteilungsaufgabe wurden noch einmal die an erster und an dritter Stelle ausgewählten Produktbeschreibungen präsentiert.[254] Die Aufgabe der Probanden war, den Präferenzunterschied zwischen diesen beiden Alternativen mit einem DM-Betrag zu bewerten. Auf Basis dieser Anwort war es möglich, die Präferenzunterschiede des gesamten Choice-Sets in eine monetäre Größe zu transformieren. Die Quantifizierung der Präferenzunterschiede zwischen den Alternativen wurde auch erfragt, wenn sich die Probanden zunächst für die Nicht-Kauf-Alternative entschieden hatten. Die Bewertung fand dann unter der Annahme statt, daß keine anderen Produktalternativen zur Verfügung stehen.

Um die Präferenzwerte aller Stimuli (über die Choice-Sets hinweg) miteinander vergleichen zu können, wurden die einzelnen Choice-Sets über die Zahlungsbereitschaft für die in jedem Choice-Set am schlechtesten bewertete Produktalternative miteinander verankert.[255] Die Berechnung der Zahlungsbereitschaft erfolgte über die Self-Explicated-Teilnutzenwerte für die Eigenschaft „Preis".[256] Einen Sonderfall hierbei stellten die Choice-Sets dar, bei denen in der ersten Auswahlentscheidung die Nicht-Kauf-Alternative gewählt wurde. In der vorliegenden Untersuchung trat dieser Fall bei 24 von insgesamt 941 bewerteten Choice-Sets und damit sehr selten auf. Für die überraschend seltene Wahl der Nicht-Kauf-Alternative können unterschiedliche Gründe eine Rolle spielen. Zum einen kann das Ergebnis bedeuten, daß überwiegend attraktive Produktprofile zur Auswahl standen, bei denen nur in den seltensten Fällen alle abzulehnen waren.

---

254 Die Auswahl genau dieser beiden Alternativen ist nicht zwingend erforderlich. Der dritte Teil der Beurteilungsaufgabe kann auch für alternative Paare aus dem Choice-Set durchgeführt werden.

255 Auch hier gilt, daß eine Verankerung über die am schlechtesten bewertete Alternative nicht unbedingt erforderlich ist. Eine Verankerung der Choice-Sets kann prinzipiell auch über die jeweils am besten bewertete oder eine „mittlere" Alternative vorgenommen werden. Empfohlen wird dennoch eine Verankerung über einen der beiden Eckwerte, d.h. die beste oder schlechteste Alternative, denn der vorliegende Datensatz belegt, daß hier die höchste Übereinstimmung zwischen den auf Basis der Self-Explicated-Daten ermittelten und im Conjoint-Teil beobachteten Präferenzrangfolgen auftrat.

256 Zur genauen Vorgehensweise der Transformation der Präferenzwerte in eine metrische Gesamtreihenfolge vgl. das Rechenbeispiel in Kapitel 3.2.2.3.

Zum anderen könnten die Probanden bei einer Befragung hinsichtlich ihrer Präferenzen zwischen verschiedenen Produkten zögern, tatsächlich alle Alternativen abzulehnen. Letzteres würde allerdings einen wesentlichen Vorteil aller Choice-Experimente in Frage stellen. Ein abschließendes Urteil diesbezüglich ist auf Basis der vorliegenden Studie jedoch nicht zulässig. Denn hierbei muß berücksichtigt werden, daß vor allem die zusätzliche Bewertung der Stimuli nach der Entscheidung für die Nicht-Kauf-Alternative die vorliegenden Ergebnisse verzerrt haben kann. Weiterhin ist nicht auszuschließen, daß die aus Platzgründen wesentlich kleinere Darstellung der Nicht-Kauf-Alternative auf dem Bildschirm gegenüber den Produktbeschreibungen dieses Ergebnis beeinflußt haben könnte.[257] Eine genauere Überprüfung der 24 Choice-Sets ergab, daß bei 15 dieser Sets auf Basis der Self-Explicated-Teilnutzen die niedrigsten Zahlungsbereitschaften für die Produktalternativen prognostiziert wurden. Konsistent mit dem geäußerten Nicht-Kauf-Interesse wurden sie durch die Verankerung der Choice-Sets über die Zahlungsbereitschaften am Ende der Präferenzrangfolge eingeordnet. Widersprüche zwischen geäußerter Nicht-Kauf-Wahl und geschätzter Zahlungsbereitschaft zeigten sich in weniger als 1% aller bewerteten Choice-Sets und können deshalb vernachlässigt werden.

Die ermittelte Gesamtpräferenzreihenfolge über alle bewerteten Stimuli bildete die abhängige Variable für eine OLS-Schätzung, um die Teilnutzenwerte für Kerneigenschaften zu bestimmen.[258] Als unabhängige Variablen gingen die Eigenschaftsausprägungen in die OLS-Schätzung ein.[259] Da die Auswahl der Stimuli nicht wie beispielsweise bei der CCC auf einem orthogonalen Design basierte, sondern sich am Zielkriterium der Pareto-Optimalität orientierte, mußten alle Gesamtdesigns auf Korrelationen zwischen den Eigenschaftsausprägungen überprüft werden. Für alle 234 Probanden wurden die paarweisen Korrelationen

---

257 Vgl. hierzu auch den Bildschirmausdruck dieses Befragungsschritts im Anhang 7.2.

258 Zur Berechnung der Teilnutzenwerte durch Regressionsanalyse vgl. auch Backhaus et al. (1996), S. 551 f.

259 Die Schätzung der Teilnutzenwerte wurde mit dem Conjoint-Modul in SPSS vorgenommen. Es sei an dieser Stelle darauf hingewiesen, daß SPSS dabei ebenfalls eine a priori-Ordnung (entsprechend der „constraints" beim Verfahren LINMAP) von Ausprägungen ermöglicht. Diese werden jedoch nicht in der Schätzung berücksichtigt, sondern lediglich im Falle einer Verletzung dieser Ordnung als solche ausgewiesen. Damit werden die Vorteile der Integration entsprechender a priori-Ordnungen bei der OLS-Schätzung nicht genutzt.

zwischen den Eigenschaftsausprägungen der Kerneigenschaften überprüft. Insgesamt ergaben sich 1404 Korrelationskoeffizienten, von denen 1361 (≈ 97%) in ihrem Betrag nicht größer als 0,3 ausfielen.[260] Sie liegen damit deutlich unter dem bei Conjoint-Designs üblicherweise angenommenen Grenzwert von 0,5.[261] Die verbleibenden 3% der paarweisen Korrelationen wiesen mit Werten zwischen 0,4 und 0,7 deutlich stärkere Abhängigkeiten auf. Aufgrund ihres sehr seltenen Auftretens muß jedoch nicht angenommen werden, daß das Gesamtergebnis, insbesondere bei Auswertung der Daten auf aggregierter Ebene, dadurch maßgeblich beeinflußt wird. Die Überprüfung der Designs auf Korrelationen zwischen den Eigenschaftsausprägungen zeigt demnach, daß insgesamt sehr geringe paarweise Korrelationen zwischen den Eigenschaftsausprägungen auftreten, diese aber zugunsten der Pareto-Optimalität der Designs in Kauf genommen werden können. Eine Umskalierung der geschätzten Conjoint-Teilnutzenwerte für die Kerneigenschaften erfolgte entsprechend dem Vorgehen bei der CCC derart, daß die Summe der Bedeutungsgewichte der Kerneigenschaften gleich der Summe der Self-Explicated-Teilnutzenwerte für diese Eigenschaften war.

ACA

Für die Gestaltung des gesamten ACA-Interviews wurden die Standardeinstellungen des Systems übernommen bzw. die Entscheidungen unter dem Gesichtspunkt einer bestmöglichen Vergleichbarkeit zu den beiden neuen Verfahren CCC und CHIC getroffen. So bildete das System 4.0 der ACA bei gegebener Anzahl an Eigenschaften und Ausprägungen 19 Paarvergleiche für die Conjoint-Phase der Befragung. Die Produktkonzepte wurden entsprechend der Anwendungsempfehlung zunächst anhand von 2 und im späteren Interviewverlauf von 3 Eigenschaftsausprägungen charakterisiert.[262] Auf einer 9-Punkte-Rating-Skala gaben die Probanden für jeden Paarvergleich an, welches der beiden Produktkonzepte

---

260 Die Gesamtzahl der paarweisen Korrelationen ergibt sich aus (n(n-1)/2)*234, wobei n für die Anzahl der Kerneigenschaften steht.

261 Es sei darauf hingewiesen, daß dieser Grenzwert als „Daumenregel" zu interpretieren ist. Eine mögliche empirische Überprüfung des Einflusses der Korrelationen auf die Schätzergebnisse besteht in einer erneuten Schätzung der Parameter, nachdem eine der beiden korrelierten Variablen ausgeschlossen wurde (vgl. Backhaus et al. (1996), S. 34). Im Rahmen einer Conjoint-Analyse ist diese Vorgehensweise jedoch nicht sinnvoll, da damit eine zu untersuchende Eigenschaft, die als unabhängige Variable in die Schätzgleichung eingeht, aus der Analyse ausgeschlossen werden müßte.

262 Vgl. Johnson (1987), S. 261 und Green/Krieger/Agarwal (1991), S. 216.

sie präferieren und wie stark.[263] Nach jedem Paarvergleich aktualisiert die ACA die Schätzergebnisse auf Basis der neu gewonnenen Informationen. So liegen nach Abschluß der Paarvergleichsaufgabe Teilnutzenwerte vor, die ausgehend von den Self-Explicated-Teilnutzenwerten (SE-ACA) schrittweise mit jedem bewerteten Paarvergleich aktualisiert wurden. Die Vorgehensweise zur Schätzung der Teilnutzenwerte unterscheidet sich infolgedessen maßgeblich von der CCC und CHIC, bei denen die Conjoint-Teilnutzenwerte auf einer einmaligen Auswertung aller erhobenen Präferenzurteile beruhen. In diesem Zusammenhang wird noch einmal deutlich, daß die ACA im Gegensatz zu den beiden neuen Verfahren nicht nur hinsichtlich der Gestaltung nachfolgender Befragungsschritte, sondern auch der Schätzung der Teilnutzenwerte adaptiven Charakter aufweist.

4.3.3 Kombinationsphase

Bei allen drei Verfahren wurden in der Kombinationsphase zusätzliche Präferenzurteile für ein Kalibrierungsset[264] erhoben, um die Informationen aus der Self-Explicated-Phase und der Conjoint-Phase optimal kombinieren zu können. Bei der CCC präsentierte das Programm hierfür 9 Produktprofile, deren Konstruktion wie in der Conjoint-Phase einem orthogonalen Designplan entsprach. Die Ausprägungen der 4 Kerneigenschaften wurden jedoch neu kombiniert, so daß sich die Produktprofile so weit wie möglich von den bereits abgefragten unterschieden. In der Kombinationsphase der CHIC hatten die Probanden jeweils ein zusätzliches Choice-Set zu beurteilen, das nicht in die Schätzung der Conjoint-Teilnutzenwerte einging. Die prinzipielle Aufgabenstellung für das Kalibrierungsset stimmte bei beiden Verfahren jeweils mit der Conjoint-Phase überein. Im Gegensatz dazu waren die Probanden in der Kombinationsphase bei der ACA aufgefordert, separat für 6 Produktkonzepte (beschrieben durch 5 Eigenschaften) eine Kaufwahrscheinlichkeit von 0% bis 100% anzugeben.[265] Dabei werden diese

---

263 Mit der Wahl einer 9-Punkte-Skala wird bei der Paarvergleichsaufgabe die gleiche Rating-Skala verwendet wie in der Self-Explicated-Phase des ACA-Interviews, was die Vergleichbarkeit der Ergebnisse zwischen diesen beiden Phasen erhöht.

264 In Zusammenhang mit der Kombinationsphase der drei Verfahren werden die Begriffe Kalibrierungsset und Validation-Set (vgl. Kapitel 3) synonym verwendet.

265 Das System der ACA 4.0 ermöglicht die Wahl zwischen mindestens 4 und maximal 9 Kalibrierungskonzepten. Sofern die Ergebnisse des Self-Explicated-Teils und der Paarvergleichsaufgabe durch optimale Gewichtung (anstatt einer Gleichgewichtung) kombiniert werden sollen, wird die Abfrage von wenigstens 5 Konzepten empfohlen (vgl. Metegrano (1994), S. 3-34). Mit der Festlegung von 6 Produktkonzepten wird diese Mindestanforderung auf jeden Fall erfüllt.

Produktkonzepte immer so konstruiert, daß zunächst das Konzept mit dem niedrigsten Nutzen und als zweites das mit dem höchsten Nutzen zu bewerten ist. Diese beiden Eckpunkte sollen als Anker für die Bewertung dazwischen liegender Produktkonzepte fungieren.[266] Für deren Konstruktion greift das Programm auf die Informationen sowohl aus der Self-Explicated-Phase als auch aus der Paarvergleichsaufgabe zurück.

Mit dieser zusätzlichen Anpassung der Kombinationsphase an die Ergebnisse der Conjoint-Phase, ist die beschriebene Reihenfolge der 3 Interviewphasen für die ACA zwingend vorgegeben. Dagegen besteht bei der CCC und der CHIC die Möglichkeit, die Abfrage des Kalibrierungssets zwischen der Self-Explicated- und der Conjoint-Phase beurteilen zu lassen.[267] Für ein solches Vorgehen spricht, daß die Gefahr reduziert wird, die Prognoseergebnisse durch eine Wiederholung der Conjoint-Aufgabe und damit verbundener Lerneffekte zu verzerren.[268] So argumentieren Huber et al. (1993) beispielsweise, daß die Validität einer Voll-Profil Conjoint-Analyse erhöht werden kann, wenn vor der eigentlichen Hauptaufgabe eine „Warmup-Phase" integriert wird, um die Probanden mit der Aufgabe vertraut zu machen.[269] Für die CCC und die CHIC besteht hierin eine Gefahr. Die Kombination der Teilnutzenwerte kann durch die Ähnlichkeit der Aufgabenstellung zwischen Conjoint- und Kombinationsphase durch einen Reihenfolgeeffekt zugunsten der Conjoint-Werte verstärkt werden. Srinivasan/Park (1997) berichten beispielsweise, daß bei 73% der Probanden die Conjoint-Teilnutzen ein höheres Gewicht erhalten als die Self-Explicated-Werte. Als einen Grund für dieses Ergebnis führen die Autoren die Ähnlichkeit zwischen der Conjoint- und Kombinationsaufgabe an.[270] Vor diesem Hintergrund diente im vorliegenden Datensatz (analog zum Vorgehen von Srinivasan/Park (1997)) jeweils das erste Design bei der CCC bzw. das erste Choice-Set bei der CHIC als Kalibrierungs-

---

266 Vgl. Metegrano (1994), S. 3-9.

267 In diesem Fall dienen die in der dritten Interviewphase (bisher Kombinationsphase) erhobenen Daten zur Schätzung von Teilnutzenwerten, die dann zur Prognose der Präferenzurteile der zweiten Interviewphase (bisher Conjoint-Phase) herangezogen werden.

268 Vgl. Srinivasan/Park (1997), S. 288.

269 Vgl. Huber et al. (1993), S. 111.

270 Vgl. Srinivasan/Park (1997), S. 289.

set.[271] Dieser Reihenfolgeeffekt spielt bei der ACA keine (wesentliche) Rolle, da sich die Aufgabe der Kombinationsphase maßgeblich von der Paarvergleichsaufgabe unterscheidet.

Die Bestimmung der optimalen Gewichtung von Self-Explicated- und der Conjoint-Teilnutzenwerte entsprach bei der CCC und der CHIC der in Kapitel 3.1 beschriebenen Vorgehensweise. Für jede Testperson wurde aus alternativen Gewichten (w_i=0; 0,01; 0,02; ... ; 1, wobei $i \in I$) jeweils das optimale Gewicht w_i* ausgewählt, bei der sich die maximale Übereinstimmung zwischen prognostizierten und beobachteten Präferenzen für das Kalibrierungsset ergab. Während bei der CCC die prognostizierten Präferenzen (gemessen durch die Gesamtnutzenwerte) in eine Rangfolge transformiert wurden, gingen bei der CHIC auch die Präferenzunterschiede in die Prognose ein. In Abhängigkeit des damit verbundenen Skalenniveaus konnte die Übereinstimmung durch den Korrelationskoeffizienten nach Kendall (bei der CCC) bzw. Pearson (bei der CHIC) gemessen werden. In Tabelle 4-4 ist eine Häufigkeitsverteilung der ermittelten Gewichtungsfaktoren w_i* für die CCC und die CHIC angegeben. Darüber hinaus sind die entsprechenden Werte in der Studie von Srinivasan/Park (1997) angeführt. Ein Vergleich der Häufigkeiten zeigt, daß ähnlich zu den Befunden von Srinivasan/Park (1997) auch bei der CCC bei mehr als der Hälfte der Probanden (60%) die Conjoint-Teilnutzenwerte ein höheres Gewicht als die der Self-Explicated-Phase aufweisen. Im Gegensatz dazu sinkt dieser Anteil bei der CHIC auf 46%. Dieser Rückgang erscheint plausibel, wenn man berücksichtigt, daß bei der CHIC die Informationen der Self-Explicated-Phase zusätzlich bei der Bestimmung der abhängigen Variable für die Conjoint-Analyse einfließen und damit implizit eine höhere Bedeutung für die Conjoint-Phase erlangen. Als zweites Ergebnis ist anzuführen, daß die beiden Randbereiche (0 bis 0,1) und (0,91 bis 1,0) bei der vorliegenden Studie stärker besetzt sind als bei den Vergleichswerten von Srinivasan/Park (1997). Somit tritt bei der CCC und der CHIC häufiger der Fall auf, daß die beste Prognose für das Kalibrierungsset mit den Ergebnissen einer der beiden Phasen (nahezu) allein erzielt werden als über deren Kombination. Ob dies insgesamt häufiger mit der Self-Explicated-Phase oder der Conjoint-Phase gelingt,

---

[271] Bei der CHIC mußte bei einem Teil der Probanden mit einem 3*3*2*2-Design der Kerneigenschaften das zweite Choice-Set als Kalibrierungsset herangezogen werden, da andernfalls die verbleibenden 8 Stimuli ein unlösbares Gesamtdesign gebildet hätten.

kann nicht eindeutig abgeleitet werden. Während bei der CCC hohe Gewichtungsfaktoren häufiger auftreten (13,2% vs. 23,1%), ist dies bei der CHIC genau umgekehrt (19,7% vs. 12,0%).

Tabelle 4-4: Häufigkeitsverteilung der optimalen Gewichtungsfaktoren w_i* im Vergleich

Optimaler Gewichtungsfaktor w_i*	**Anteil der Probanden in %**					
	CCA[a)]		**CCC**		**CHIC**	
0 bis 0,1[b)]	2,8		13,2		19,7	
0,11 bis 0,2	3,7		6,6		9,8	
0,21 bis 0,3	3,7		7,0		9,8	
0,31 bis 0,4	5,6		4,1		9,0	
0,41 bis 0,5	11,1		8,7		5,6	
0,51 bis 0,6	9,3	≅73	7,4	≅60	13,7	≅46
0,61 bis 0,7	14,8		6,6		8,5	
0,71 bis 0,8	20,4		9,1		7,3	
0,81 bis 0,9	19,4		14,0		4,7	
0,91 bis 1,0	9,2		23,1		12,0	
Gesamt:	100,0		100,0		100,0	

a) Quelle: Srinivasan/Park (1997), S. 289

b) Ein Gewichtungsfaktor z.B. von 0,1 impliziert, daß der Conjoint-Teilnutzen mit einem Gewicht von 0,1 und der Self-Explicated-Teilnutzen mit einem Gewicht von 0,9 in die kombinierten Teilnutzenwerte eingeht (vgl. Gleichung (3-2) in Kapitel 3.1.3).

Bei der ACA wird die Gewichtung durch eine Regressionsanalyse ermittelt, bei der die Logits der erhobenen Kaufwahrscheinlichkeiten für die Kalibrierungskonzepte vorhergesagt werden (vgl. Abschnitt 3.3.3). Als Ergebnis liefert das System neben einer Konstanten zwei Regressionskoeffizienten, mit denen die Self-Explicated-Teilnutzenwerte und die der Paarvergleichsaufgabe gewichtet werden, um die Logits der erhobenen Kaufwahrscheinlichkeiten bestmöglich prognostizieren zu können. Diese Regressionskoeffizienten entsprechen inhaltlich den Gewichtungsfaktoren w_i* bei der CCC und der CHIC. Da die Regressionskoeffizienten nicht wie die Gewichtungsfaktoren w_i in einem festen Wertebereich definiert sind, ist eine direkte Gegenüberstellung in ihrer absoluten Höhe nicht möglich. Aus diesem Grund sind die Regressionskoeffizienten der ACA

nicht in Tabelle 4-4 ausgewiesen. Ein Vergleich zwischen den Regressionskoeffizienten für den Self-Explicated-Teil und die Paarvergleichsaufgabe innerhalb der ACA ist dagegen zulässig. Eine entsprechende Auswertung dieser Koeffizienten ergab, daß bei 72 der 249 Probanden, also in 29 % der Fälle, die geschätzten Teilnutzenwerte der Paarvergleichsaufgabe mit einem höheren Gewicht als die Self-Explicated-Teilnutzenwerte eingehen. Dieser Anteil fällt deutlich geringer aus als bei der CCC und CHIC. Zieht man zudem in Betracht, daß bei weiteren 160 Probanden (also fast 65% aller Probanden) die Teilnutzenwerte exakt gleich gewichtet wurden, bestätigt dies implizit die Vermutung, daß durch Wiederholung der Conjoint-Aufgabe in der Kombinationsphase die Ergebnisse bzw. die Gewichtung der Teilnutzenwerte zugunsten der wiederholten Aufgabe verzerrt werden können. Der Anteil an Probanden, bei denen $w_i^* > 0{,}5$ ist und damit den Conjoint-Teilnutzen ein höheres Gewicht als den Self-Explicated-Teilnutzen zukommt, ist bei beiden neuen Verfahren deutlich höher als bei der ACA (vgl. hierzu die Häufigkeiten für $w_i^* > 0{,}5$ bei der CCC und CHIC in Tabelle 4-4). Gleichzeitig ist durch die Gegenüberstellung der Ergebnisse in Tabelle 4-4 ersichtlich, daß dieser Effekt in der vorliegenden Untersuchung weniger stark ausgeprägt ist als bei Srinivasan/Park (1997).

4.3.4 Zusatzdaten und Kaufsimulation

In Ergänzung zum eigentlichen Conjoint-Interview wurden jeweils am Ende der Befragungen demografische Daten sowie Informationen zu Kauf- und Markenerfahrung bei Kühlschränken bzw. Haushaltsgroßgeräten erhoben. Den Abschluß des Gesamtinterviews bildete eine simulierte Kaufentscheidung. Dieser Befragungsschritt diente als Holdout-Set, um die Prognosegüte der alternativen Verfahren überprüfen zu können. Die Validierung alternativer Methoden zur Präferenzmessung auf Basis einer Holdout-Aufgabe ist in der Literatur sehr weit verbreitet und liefert die Grundlage der meisten Validitätsvergleiche.[272]

Trotz der häufigen Anwendung einer Holdout-Aufgabe zur Validierung der Ergebnisse ist die Frage nach der Gestaltung eines Holdout-Sets in der Literatur bisher vernachlässigt worden. Die Entscheidungshilfen zur Zusammenstellung der Holdout-Konzepte beschränken sich bislang darauf, daß zum einen keine do-

---

[272] Vgl. hierzu die Dominanz dieses Kriteriums im Literaturüberblick in Kapitel 2 der vorliegenden Arbeit.

minanten Alternativen im Set enthalten sein sollen. Zum anderen ist darauf zu achten, daß die Angebote des Holdout-Sets in ihrer Attraktivität nicht zu ähnlich sind. Die Einhaltung dieser beiden Prinzipien wird allerdings dadurch erschwert, daß oft keine a priori Kenntnisse über die Präferenzstrukturen der Probanden vorliegen. In diesem Fall können keine Annahmen hinsichtlich der Attraktivität alternativer Produktkonzepte getroffen und bei der Konzeption des Holdout-Sets berücksichtigt werden.[273] Einen Lösungsvorschlag hierzu liefern Herrmann/Schmidt-Gallas/Huber (2000), die das Holdout-Set im Anschluß an ein ACA-Interview ebenfalls computergestützt und unter Berücksichtigung der bereits erhobenen Informationen für alle Probanden individuell gestalten.[274] Wittink/Bergestuen (1999) führen eine dritte Bedingung an. Die Holdout-Aufgabe sollte so nah wie möglich an der tatsächlichen Kaufsituation orientiert sein.[275] Aufgrund dessen bestand das Holdout-Set aus 5 Produktbeschreibungen, die zum Zeitpunkt der Befragung real am Markt existierenden Angeboten des betrachteten Preissegments für Tischkühlschränke entsprachen. Um zu vermeiden, daß die Produktkonzepte des Holdout-Sets zu ähnlich sind, wurde die gesamte Bandbreite aller betrachteten Eigenschaftsausprägungen berücksichtigt. Insgesamt wurden 5 Eigenschaften zur Beschreibung der Kühlschränke herangezogen.[276] Da insbesondere für die Eigenschaft „Marke" keine a priori Informationen über die Präferenzen bei den Probanden vorliegen, kann über die Dominanz der verschiedenen Alternativen keine eindeutige Aussage gemacht werden.[277] Tabelle 4-5 zeigt die Produktkonzepte der Holdout-Aufgabe.

---

273 Zu den Empfehlungen bezüglich der Gestaltung eines Holdout-Sets vgl. Johnson (1997a).

274 Vgl. Herrmann/Schmidt-Gallas/Huber (2000), S. 264 ff.

275 Vgl. Wittink/Bergestuen (1999), S. 10.

276 Es handelte sich um die Eigenschaften Marke, Preis, Energieverbrauch, Nutzinhalt und Sterne des Gefrierfachs. Es sind dies die Eigenschaften, deren Ausweis bei der Produktpräsentation gesetzlich vorgeschrieben ist.

277 Ähnliches gilt für die Eigenschaft „Sterne des Gefrierfachs", da ein 3-Sterne-Fach gegenüber einem 1-Sterne-Fach nicht eindeutig präferiert wird.

Tabelle 4-5: Produktbeschreibungen im Holdout-Set für die simulierte Kaufentscheidung

Angebot 1	Angebot 2	Angebot 3	Angebot 4	Angebot 5
AEG	Liebherr	Siemens	Electrolux	Quelle Privileg
145 Liter	145 Liter	125 Liter	120 Liter	120 Liter
***-Fach	*-Fach	***-Fach	***-Fach	***-Fach
208 kWh/Jahr	222 kWh/Jahr	245 kWh/Jahr	237 kWh/Jahr	175 kWh/Jahr
659,- DM	449,- DM	599,- DM	549,- DM	649,- DM

Die 5 Produktbeschreibungen wurden den Probanden am Ende des Interviews auf einem schriftlichen Fragebogen präsentiert. Die Beurteilungsaufgabe für alle Probanden bestand darin, für die Angebote des Holdout-Sets die Rangfolge anzugeben, in der sie diese Kühlschränke am ehesten kaufen würden.

4.3.5 EDV-Implementierung

Bei den in dieser Arbeit betrachteten Verfahren handelt es sich um vollständig PC-gestützte Interviews. Die Verfahren CCC und CHIC sind Windows-gestützte Befragungsprogramme, die mit der objektorientierten Programmiersprache Delphi 1.0 programmiert wurden. Die erstmalige Entwicklung der für die Befragung eingesetzten Software beanspruchte einen erheblichen Zeit- und Kostenaufwand. So fielen allein für die Programmierarbeiten bei der Entwicklung der CCC und CHIC mehr als 1000 Personalstunden an.[278] Neben der eigentlichen Conjoint-Aufgabe wurden auch sämtliche sozio-demografischen Daten und Zusatzfragen zu Kauferfahrungen im Rahmen des PC-Interviews erfaßt. Für das ACA-Interview wurde die Version 4.0 von Sawtooth Software eingesetzt.[279] Es handelt sich dabei um eine DOS-Version, die sowohl die Datenerhebung als auch die direkte Datenauswertung und verschiedene Simulationsmodule umfaßt.[280] Durch eine Verbindung des ACA-Interviews mit dem Befragungsprogramm Ci3, eben-

---

[278] Für die Unterstützung bei den Programmierarbeiten dankt die Autorin ganz herzlich Frau Simone Martin.

[279] Für die Bereitstellung des Programms sei dem Marktforschungsinstitut IPSOS Deutschland an dieser Stelle ausdrücklich gedankt.

[280] Mittlerweile liegt mit Sensus-TradeOff 1.0 und 2.0 eine Windows-gestützte Version der ACA vor. Ein Einsatz dieser multimedialen Version war zum Zeitpunkt der Datenerhebung leider noch nicht möglich.

falls von Sawtooth Software, konnten auch hier alle demografischen Daten und Zusatzfragen PC-gestützt erhoben werden. Im Anhang der Arbeit sind exemplarisch die Bildschirmmasken der einzelnen Befragungsschritte bei den neu entwickelten Verfahren angeführt.

Mit der Integration der Datenauswertung in den Interviewablauf bei der ACA wird nicht nur eine stärkere Anpassung der Befragungsschritte an vorangegangene Antworten ermöglicht, sondern auch ein entscheidender Vorteil unter Anwendungsgesichtspunkten gegenüber den beiden neu vorgestellten Verfahren geboten. Innerhalb eines Softwarepaketes liefert die ACA dem Marktforscher das Befragungsprogramm und für jede Interviewphase die geschätzten Teilnutzenwerte. Darüber hinaus bietet das System die Möglichkeit, verschiedene Simulationsanalysen durchzuführen. Im Gegensatz hierzu liegen nach Abschluß der CCC- und CHIC-Interviews zunächst allein die „Rohdaten" vor, die in weiteren Analyseschritten vom Marktforscher ausgewertet werden müssen. Die vorgenommene Individualisierung der Befragungsschritte und speziell der Designerstellung bei der Conjoint-Aufgabe erfordern dabei zunächst eine separate Auswertung der erhobenen Daten jedes einzelnen Probanden. Um den neuen Verfahren eine Verbreitung in der Praxis zu ermöglichen, ist (zumindest teilweise) die Implementierung der Datenauswertung in einem Software-Paket unerläßlich und bereits Schwerpunkt weiterer Forschungsprojekte.

5 Empirischer Validitätsvergleich

In Kapitel 3 sind die beiden neuen adaptiven Verfahren CCC und CHIC sowie die ACA vorgestellt worden. Neben der Darstellung der einzelnen Befragungsschritte in der Self-Explicated-, Conjoint- und Kombinationsphase wurden die methodischen Unterschiede zwischen diesen Verfahren aufgezeigt. Nach dieser konzeptionellen Gegenüberstellung werden sie im folgenden einer empirischen Validitätsprüfung unterzogen. Die ACA als etabliertes Verfahren dient dabei als Referenzpunkt für die Ergebnisgüte der beiden neuen Methoden. Gleichzeitig kann ein direkter Vergleich der neuen Verfahren mit der ACA zeigen, ob alternative Vorgehensweise vor allem bei der Datenerhebung in der Conjoint-Phase zu valideren Ergebnissen führen. Beispielsweise ist bei der ACA bisher die Konzeption der Paarvergleichsaufgabe häufig als Schwachpunkt des Verfahrens angeführt worden.[281]

Ein wichtiges Ergebnis der Zusammenstellung bisheriger empirischer Vergleichsstudien war das relativ gute Abschneiden der Self-Explicated-Methode sowohl im Vergleich zu rein dekompositionellen als auch zu unterschiedlichen hybriden Varianten der Conjoint-Analyse. Zahlreiche Vergleiche bescheinigten der Self-Explicated-Methode eine gleich gute und teilweise sogar eine signifikant höhere Validität.[282] Vor dem Hintergrund dieser Ergebnisse wird die Self-Explicated-Methode als alternative Vorgehensweise zur Präferenzmessung auch in den folgenden Validitätsvergleich einbezogen. Hierfür wird bei allen drei adaptiven Verfahren ACA, CCC und CHIC eine separate Auswertung der ersten Interviewphase (Self-Explicated-Phase) vorgenommen und mit SE-ACA, SE-CCC und SE-CHIC bezeichnet. Der Validitätsvergleich umfaßt somit insgesamt 8 unterschiedliche Ansätze. Es sind dies die drei Self-Explicated-Varianten[283] sowie die drei adaptiven Verfahren ACA, CCC und CHIC, wobei bei den neuen Verfahren jeweils zwischen CCC-alt und CCC-neu bzw. CHIC-alt und CHIC-neu

---

281 Vgl. Green/Srinivasan (1990), S. 11; Green/Krieger/Agarwal (1991), S. 220 f.; Reiners (1996), S. 78 f.; Srinivasan/Park (1997), S. 286.

282 Vgl. hierzu die Ausführungen in Kapitel 2.3.2 und 2.5.2.

283 Dabei sei einschränkend angemerkt, daß nur 2 verschiedene Varianten der Self-Explicated-Methode verglichen werden, da die Vorgehensweise bei der SE-CCC und SE-CHIC identisch ist.

differenziert wird, um den Effekt der modifizierten Schätzmethode in der Kombinationsphase quantifizieren zu können (vgl. Abschnitt 3.1.3 und 3.2.3).

Die vergleichende Validitätsprüfung aller 8 Methoden wird anhand verschiedener Gütekriterien vorgenommen. Es sind dies die gleichen Kriterien, die bereits zur Systematisierung der bisherigen Forschungsergebnisse in Kapitel 2 dienten:

1. diskriminierende Validität,
2. interne Validität,
3. Prognosevalidität,
4. externe Validität.[284]

Einzige Ausnahme bildet die Reliabilität, die in der vorliegenden Studie nicht als weiteres Gütekriterium integriert wurde. Voraussetzung für einen entsprechenden Test wäre eine zumindest teilweise Wiederholung der Befragung derselben Person mit der gleichen Datenerhebungsmethode. Auf diesen zusätzlichen Befragungsschritt wurde verzichtet, um die Auskunftsbereitschaft der Probanden nicht zu überfordern. Die Ergebnisse hinsichtlich der vier Gütekriterien werden in den folgenden vier Kapiteln (5.1 - 5.4) dargestellt. Abschnitt 5.5 faßt die wesentlichen Befunde der empirischen Validitätsprüfung noch einmal zusammen.

5.1 Diskriminierende Validität

Mit dem Kriterium der diskriminierenden Validität soll überprüft werden, ob die untersuchten Methoden zur Messung von Nachfragerpräferenzen zu unterschiedlichen Ergebnissen in bezug auf die geschätzten Präferenzstrukturen führen. Da sich die betrachteten Verfahren in der Datenerhebungsmethode und/oder in der Vorgehensweise zur Schätzung der Teilnutzenwerte z.T. maßgeblich unterscheiden, ist zu vermuten, daß die ermittelten Präferenzstrukturen in Abhängigkeit von der eingesetzten Methode variieren.[285] Somit gilt es zu testen, ob und inwieweit die Ergebnisse von der Vorgehensweise zur Erfassung der Präferenzstrukturen beeinflußt werden. Treten in diesem Punkt keine Unterschiede zwischen den Verfahren auf, kann für alle Methoden die gleiche Prognosegüte erwartet werden.

---

[284] Zur Definition und Abgrenzung der Kriterien vgl. Kapitel 2.2.2.1.

[285] Es sei darauf hingewiesen, daß an dieser Stelle ebenso die konvergierende Validität getestet werden könnte. Wie in Kapitel 2.2.2.1 diskutiert, entspricht eine Überprüfung der konvergierenden Validität der gleichen Vorgehensweise, wobei lediglich die Ausgangshypothese für die Tests umgekehrt formuliert wird. In diesem Fall werden *keine* Unterschiede zwischen den verschiedenen Verfahren erwartet.

Diese Folgerung begründet sich darin, daß alle weiteren Analysen auf den ermittelten Präferenzstrukturen basieren, die durch die Teilnutzenwerte und Bedeutungsgewichte abgebildet werden. Für die Anwendung der Verfahren in der Marktforschung würde ein solches Ergebnis bedeuten, daß nur Kosten- und Zeitaufwand beim Einsatz eines Verfahren ausschlaggebende Entscheidungskriterien sein können, ohne die Verläßlichkeit und den Aussagegehalt der Ergebnisse einzuschränken.

5.1.1 Diskriminierende Validität zwischen alternativen Formen der Präferenzabfrage

In Kapitel 2 wurden die dekompositionellen Ansätze zur Präferenzmessung anhand der Kriterien „Anzahl simultan zu bewertender Stimuli" und „Gestaltung der Präferenzabfrage bei der Bewertung der Produktkonzepte" abgegrenzt. Während die Anzahl der Stimuli in der vorliegenden Untersuchung nicht unabhängig von weiteren Einflußgrößen variiert wurde, kann der Effekt der Präferenzabfrage im Rahmen des CCC-Interviews isoliert betrachtet werden. In der Conjoint-Phase der CCC wurden die Präferenzen für die verschiedenen Produktkonzepte sowohl durch Rangreihung als auch anhand einer Kaufwahrscheinlichkeitsskala von 0% bis 100% erhoben. Der vorliegende Datensatz bietet durch diese Zweistufigkeit der Beurteilungsaufgabe eine direkte Vergleichsmöglichkeit zwischen den beiden Formen der Präferenzabfrage. Eine Gegenüberstellung der Ergebnisse wird zeigen, ob die zusätzliche Quantifizierung der Präferenzunterschiede in Form von Kaufwahrscheinlichkeiten zu unterschiedlichen, im besten Fall valideren Ergebnissen führt.

Für den Vergleich der Präferenzstrukturen wurden die individuellen Bedeutungsgewichte für die Kerneigenschaften sowohl auf Basis der Rangdaten als auch unter Verwendung der Kaufwahrscheinlichkeitsdaten berechnet. Grundlage hierfür bilden jeweils die geschätzten Teilnutzenwerte. Dividiert man die Spannweite[286] der Teilnutzenwerte einer Eigenschaft durch die Summe aller Spannweiten der betrachteten Eigenschaften, erhält man die relative Wichtigkeit dieser Eigenschaft, die üblicherweise als Bedeutungsgewicht bezeichnet wird.[287] Aufgrund

---

[286] Die Spannweite stellt den Unterschied zwischen maximalem und minimalem Teilnutzenwert einer Eigenschaft dar.

[287] Vgl. z.B. Mengen/Simon (1996), S. 234.

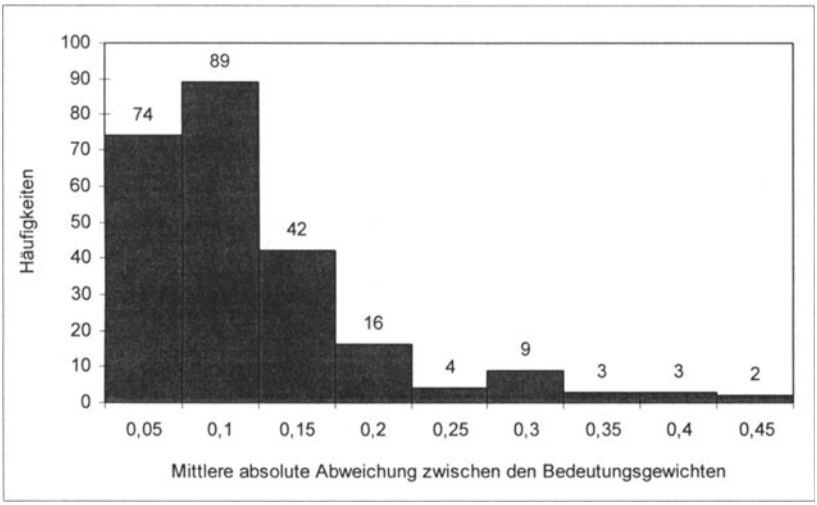
des unterschiedlichen Skalenniveaus der erhobenen Präferenzdaten erfolgte die Bestimmung der Teilnutzenwerte für die betrachteten Fälle nicht durch das gleiche Schätzverfahren. So wurden die Rangdaten mit LINMAP und die Kaufwahrscheinlichkeitsdaten mittels einer OLS-Schätzung ausgewertet. Damit wird die isolierte Betrachtung der Präferenzabfrage teilweise eingeschränkt. Allerdings hätte der Vergleich durch Einsatz des gleichen Schätzverfahrens an Aussagegehalt verloren. Eine Auswertung der Kaufwahrscheinlichkeitsdaten mit LINMAP ist grundsätzlich möglich, da das Verfahren als abhängige Variable auch Rating-Daten zuläßt. In der Schätzung berücksichtigt LINMAP allerdings nur die ordinalen Informationen der Präferenzdaten.[288] Aus diesem Grund wurde die Schätzung der Teilnutzenwerte mittels OLS-Methode vorgenommen, wobei die Kaufwahrscheinlichkeitsdaten die abhängige Variable bildeten. Der umgekehrte Fall, also eine Auswertung der Rangdaten auf Basis einer OLS-Schätzung, würde eher eine Überprüfung des geeigneten Schätzverfahrens bei ordinalem Skalenniveau anstatt der Form der Präferenzabfrage bedeuten. Aufgrund dieser Argumente wurde der zusätzliche Einflußfaktor auf die Ergebnisse durch unterschiedliche Schätzverfahren in Kauf genommen und abhängig vom Skalenniveau der Präferenzdaten das entsprechende Schätzverfahren eingesetzt.

Als Gütemaß zur Quantifizierung der Unterschiede zwischen den beiden Formen der Beurteilungsaufgabe wurde die mittlere absolute Abweichung zwischen den individuellen Bedeutungsgewichten für alle 4 Kerneigenschaften berechnet. Eine Häufigkeitsverteilung dieses individuellen Gütemaßes über alle Probanden ist in Abbildung 5-1 grafisch dargestellt.

---

[288] Vgl. hierzu auch die Argumentation hinsichtlich des eingesetzten Schätzverfahrens bei der CCC in Kapitel 4.3.2.

Abbildung 5-1: Häufigkeitsverteilung der individuellen mittleren absoluten Abweichungen zwischen den Bedeutungsgewichten auf Basis der Rangdaten und der Kaufwahrscheinlichkeiten (n=242)



Der Mittelwert der mittleren absoluten Abweichung über alle 242 Probanden beträgt MW=0,096 (Standardabweichung SA=0,071). Da die Bedeutungsgewichte eine relative Größe darstellen, impliziert dieser Wert, daß die geschätzten Bedeutungsgewichte der individuellen Kerneigenschaften, ermittelt auf Basis der beiden alternativen Präferenzabfragen, durchschnittlich um weniger als 10% voneinander abweichen. Die Häufigkeitsverteilung verdeutlicht dabei, daß 221 Probanden um diesen Mittelwert schwanken. Lediglich 21 Probanden weisen eine mittlere absolute Abweichung auf, die größer als 0,2 ist.

Die Auswertung dieses Güteindex zeigt, daß je nach Form der Präferenzabfrage auf individueller Ebene z.T. Abweichungen in den Ergebnissen auftreten. Unklar ist allerdings, ob diese Unterschiede statistisch signifikant sind. Ein entsprechender Test auf individuellem Niveau ist aufgrund der unterschiedlichen Schätzverfahren problematisch. Um dennoch eine Aussage zur Signifikanz der Unterschiede machen zu können, wurde für jede Eigenschaft ein Test auf Verteilungsunterschiede durchgeführt. Zur Anwendung kam ein Wilcoxon-Test für Paardifferenzen bei zwei abhängigen Stichproben, wodurch der direkte Vergleich der beiden Schätzergebnisse bei jeder Testperson gewährleistet war. Es handelt sich bei diesem Test um ein verteilungsfreies Verfahren, das zwei abhängige Stichproben auf

Gleichheit hinsichtlich ihrer zentralen Tendenz überprüft.[289] Die Testergebnisse gibt Tabelle 5-1 wieder. Der Stichprobenumfang je Eigenschaft hängt davon ab, wie häufig die entsprechende Eigenschaft als Kerneigenschaft auftritt.

Tabelle 5-1: Unterschiede zwischen den Bedeutungsgewichten auf Basis der Ranking- und den Kaufwahrscheinlichkeitsdaten bei der CCC (Wilcoxon-Test für abhängige Stichproben)

Eigenschaft	**Häufigkeit als Kerneigenschaft**	**Kritische Irrtumswahrscheinlichkeit**	**Unterschied zwischen den Stichproben**
Marke	96	p=0,892	n.s.
Preis	152	p=0,365	n.s.
Energiekosten pro Jahr	192	p=0,054	schwach signifikant
Garantie	99	p=0,555	n.s.
Nutzinhalt	120	p=0,494	n.s.
Innenausstattung	70	p=0,647	n.s.
Abtauautomatik	140	p=0,045	schwach signifikant
Gefrierfach	99	p=0,205	n.s.

Das zentrale Ergebnis der Verteilungstests ist, daß für 6 der 8 Eigenschaften keine signifikanten Unterschiede (p>0,2) zwischen den Ergebnissen der beiden Präferenzabfragen auftreten. Lediglich für die Eigenschaften „Energiekosten" und „Abtauautomatik" kann die Hypothese auf Gleichheit der beiden Stichproben nicht angenommen werden. Allerdings weisen die in Tabelle 5-1 angegebenen kritischen Irrtumswahrscheinlichkeiten von p=0,054 und p=0,045 nur auf einen schwach signifikanten Unterschied hin.

Die Überprüfung der diskriminierenden Validität innerhalb der Conjoint-Phase der CCC liefert damit ein erstes wichtiges Ergebnis. Beide Formen der Präferenzabfrage führen zu vergleichbaren Ergebnissen hinsichtlich der geschätzten Präferenzstrukturen der Probanden. Ein Vergleich der Bedeutungsgewichte auf individuellem Niveau zeigt, daß diese teilweise in geringem Maße voneinander abweichen. Ein Verteilungstest belegt aber gleichzeitig, daß bei sechs Eigen-

[289] Zum Wilcoxon-Test auf Gleichheit zweier abhängiger Stichproben vgl. Bortz (1993), S. 144 f.

schaften diese Unterschiede in den Bedeutungsgewichten nicht und bei zwei Eigenschaften lediglich schwach signifikant sind. Für die vorliegende Untersuchung kann somit keine diskriminierende Validität zwischen einer Präferenzabfrage auf Basis von Rangdaten und der Abfrage von Kaufwahrscheinlichkeiten nachgewiesen werden. Dieses Ergebnis entspricht in seiner Tendenz den empirischen Befunden bisheriger Vergleiche zwischen Ranking- und Rating-Daten. Der Informationsgewinn durch Bewertung der Präferenzunterschiede hat sich bei den meisten Vergleichen nicht in den empirischen Resultaten niedergeschlagen.[290]

Unter Anwendungsgesichtspunkten impliziert dieses Ergebnis, daß eine zusätzliche Präferenzabfrage in Form von Kaufwahrscheinlichkeiten, um die Abstände zwischen den Präferenzunterschieden zu quantifizieren, nicht zwingend notwendig ist. Bereits mit weniger Informationen können vergleichbare Resultate hinsichtlich der Präferenzstrukturen erzielt werden. Ein Verzicht auf den zweiten Teilschritt der Präferenzabfrage führt zu einer Verkürzung der Interviewzeit bei der CCC und kann folglich die Kosten und den Aufwand der Befragung senken. Entsprechend wird in den folgenden Auswertungen und Validitätsprüfungen nicht mehr zwischen den beiden Teilschritten der Conjoint-Phase innerhalb der CCC differenziert, sondern nur die Ergebnisse der Rangdaten betrachtet. Selbstverständlich ist auch eine Vereinfachung der Präferenzabfrage durch Verzicht auf die Rangreihung der Produktkonzepte und der ausschließlichen Abfrage von Kaufwahrscheinlichkeiten möglich. Im vorliegenden Datensatz erfolgte die Angabe der Kaufwahrscheinlichkeiten jedoch erst im Anschluß an die vorgenommene Rangreihung und stellt in diesem Fall keine unabhängige Rating-Aufgabe dar. Aus diesem Grund fiel die Entscheidung zur Vereinfachung auf die Rangdaten.

5.1.2 Diskriminierende Validität zwischen den untersuchten Methoden zur Präferenzmessung

Mit dem vorangestellten Vergleich zwischen den Ergebnissen der beiden Formen der Präferenzabfrage wurde getestet, ob überhaupt Unterschiede auftreten. Im Gegensatz dazu werden bei einer Überprüfung der diskriminierenden Validität zwischen den untersuchten Methoden zur Präferenzmessung konkrete Unter-

---

[290] Vgl. hierzu die Ergebnisse der Studien von Huber/Herrmann/Gustafsson (2000); Leigh/MacKay/Summers (1984); Kalish/Nelson (1991) und Stallmeier (1993) in Kapitel 2.2.2.

schiede in den Ergebnissen erwartet. Wie bereits im Rahmen des theoretischen Vergleichs zwischen der Conjoint-Analyse und der Self-Explicated-Methode diskutiert, liegt ein wesentliches Problem bei letzterem Verfahren in der direkten Frage nach der Wichtigkeit der Eigenschaften.[291] Die Gefahr ist, daß Probanden das Konzept der Wichtigkeit je nach Art der Operationalisierung unterschiedlich interpretieren. In der Folge werden häufig wichtige Eigenschaften unterschätzt und weniger wichtige in ihrer Bedeutung überschätzt. Verstärkt werden kann dieser Effekt, wenn alle Eigenschaften separat zu bewerten sind. Die Probanden sind in diesem Fall nicht zu einem Trade-Off zwischen den verschiedenen Eigenschaften gezwungen, was sich in sehr ähnlichen Bedeutungsgewichten für alle bewerteten Eigenschaften widerspiegelt. Empirische Belege hierfür liefern z.B. Teas/Dellva (1985) und Stallmeier (1993). Die Autoren stellen die geschätzten Bedeutungsgewichte auf Basis einer Self-Explicated-Variante denen der Conjoint-Analyse gegenüber. In beiden Studien zeigen sich für die Conjoint-Analyse größere Unterschiede in den Bedeutungsgewichten der verschiedenen Eigenschaften als bei direkter Frage nach deren Wichtigkeit.[292] Darüber hinaus werden zwischen der SE-CCC und der SE-CHIC keinerlei Unterschiede erwartet, da es sich hierbei um identische Methoden handelt. Somit wird entsprechend der vorgenommenen Begriffsabgrenzung der Gütemaße nicht die diskriminierende, sondern die konvergierende Validität zwischen diesen beiden Verfahren überprüft.

Getestet wird die diskriminierende bzw. zwischen der SE-CCC und SE-CHIC die konvergierende Validität auch in der vorliegenden Studie im Hinblick auf die ermittelte Präferenzstruktur der Probanden. Hierfür werden separat für jede der acht Untersuchungsmethoden die individuellen Bedeutungsgewichte aller Produkteigenschaften berechnet und jeweils die Mittelwerte (vgl. Tabelle 5-2) gebildet.

---

[291] Vgl. Kapitel 2.3.1.

[292] Vgl. Teas/Dellva (1985), S. 108; Stallmeier (1993), S. 295.

Tabelle 5-2: Mittelwerte (Standardabweichungen) der Bedeutungsgewichte der Eigenschaften bei unterschiedlichen Ansätzen zur Präferenzmessung

Eigenschaft	ACA	CCC-alt	CCC-neu	CHIC-alt	CHIC-neu	SE-ACA	SE-CCC	SE-CHIC
Marke	0,111 (0,049)	0,118 (0,102)	0,125 (0,165)	0,108 (0,070)	0,104 (0,106)	0,103 (0,047)	0,102 (0,051)	0,102 (0,061)
Preis	0,117 (0,050)	0,134 (0,098)	0,146 (0,157)	0,134 (0,071)	0,145 (0,109)	0,125 (0,032)	0,135 (0,043)	0,136 (0,047)
Energiekosten pro Jahr	0,142 (0,054)	0,153 (0,104)	0,190 (0,160)	0,146 (0,082)	0,168 (0,125)	0,134 (0,030)	0,146 (0,041)	0,150 (0,044)
Garantie	0,115 (0,049)	0,106 (0,060)	0,086 (0,089)	0,126 (0,065)	0,120 (0,096)	0,130 (0,030)	0,125 (0,040)	0,127 (0,041)
Nutzinhalt	0,103 (0,043)	0,108 (0,060)	0,102 (0,101)	0,112 (0,056)	0,106 (0,079)	0,119 (0,031)	0,122 (0,040)	0,119 (0,041)
Innen-ausstattung	0,122 (0,050)	0,125 (0,070)	0,095 (0,110)	0,120 (0,054)	0,103 (0,085)	0,127 (0,031)	0,125 (0,043)	0,121 (0,043)
Abtau-automatik	0,135 (0,054)	0,113 (0,070)	0,100 (0,115)	0,113 (0,053)	0,102 (0,077)	0,131 (0,033)	0,116 (0,044)	0,118 (0,043)
Gefrierfach	0,155 (0,059)	0,144 (0,089)	0,155 (0,153)	0,141 (0,057)	0,152 (0,097)	0,131 (0,037)	0,130 (0,045)	0,128 (0,044)
Minimum	0,103	0,106	0,086	0,108	0,102	0,103	0,102	0,102
Maximum	0,155	0,153	0,190	0,146	0,168	0,134	0,146	0,150

Ein Vergleich der Ergebnisse kann die eingangs formulierte Erwartung nur teilweise bestätigen. Für fast alle Methoden zeigt sich ein verhältnismäßig geringer Unterschied zwischen dem mittleren Bedeutungsgewicht der unwichtigsten (vgl. Minimum) und der wichtigsten (vgl. Maximum) Eigenschaft. Im Durchschnitt sind alle Eigenschaften sowohl bei den adaptiven als auch bei den Self-Explicated-Verfahren annähernd gleich wichtig und schwanken zwischen ≈10% und ≈15%. Einzige Ausnahme bildet die CCC-neu, bei der die Bedeutungsgewichte insgesamt stärker variieren. Hier ist der Unterschied zwischen wichtigster (19%) und unwichtigster Eigenschaft (8,6%) mit über 10% doppelt so groß wie bei allen anderen Verfahren. Werden zudem die Standardabweichungen der dargestellten Mittelwerte in den Vergleich einbezogen, ergibt sich ein differenzierteres Bild. So weisen beide Versionen der CCC und der CHIC wesentlich höhere Standardabweichungen vom Mittelwert auf als die Self-Explicated-Verfahren und die ACA. Dies deutet auf eine größere Homogenität hinsichtlich der Präferenzstrukturen in den Stichproben der letztgenannten Verfahren hin. Für die CCC

und die CHIC dagegen heißt das, daß zumindest für eine Teilmenge der Probanden wesentlich stärkere Unterschiede zwischen der Wichtigkeit der verschiedenen Eigenschaften auftreten. Das Ergebnis, daß die ACA gegenüber der CCC und CHIC unter Berücksichtigung der Standardabweichungen geringere Unterschiede in den Bedeutungsgewichten aufweist, stimmt unterdessen mit den Befunden von Orme/Alpert/Christensen (1997) überein. Die Autoren bestätigen, daß die Wichtigkeiten der Eigenschaften bei der ACA „flacher" verlaufen als bei Voll-Profil Conjoint-Analysen.[293]

Nach diesem vorwiegend deskriptiven Vergleich der Bedeutungsgewichte sollen in einem weiteren Schritt die Unterschiede zwischen den Verfahren auf Signifikanz überprüft werden. Da für 52 der 64 (8 Eigenschaften * 8 Methoden) getesteten Verteilungen der Bedeutungsgewichte die Hypothese auf Normalverteilung abgelehnt ($p<0,1$) werden mußte, war in den meisten Fällen die Anwendungsvoraussetzung für einen T-Test auf Mittelwertunterschiede nicht erfüllt.[294] Ein Vergleich der Verteilungen der geschätzten Bedeutungsgewichte wurde deshalb anhand eines Kolmogoroff-Smirnoff-Tests vorgenommen. Dies ist ein verteilungsfreies Verfahren, das nicht an die Anwendungsvoraussetzung der Normalverteilung gebunden ist und ermöglicht, die Abweichungen in den Verteilungen der geschätzten Bedeutungsgewichte auf Signifikanz zu testen. Alle Methoden wurden jeweils paarweise und separat für jede Eigenschaft miteinander verglichen. Die Ergebnisse der insgesamt 224 durchgeführten Tests sind in zusammengefaßter Darstellung in Tabelle 5-3 wiedergegeben.[295] In der Tabelle werden aus Gründen der Übersichtlichkeit jeweils nur die Eigenschaften ausgewiesen, bei denen die Verteilungen der geschätzten Bedeutungsgewichte nicht signifikant voneinander abweichen und folglich keine diskriminierende Validität zwischen dem jeweils betrachteten Methodenpaar vorliegt.

---

293 Vgl. Orme/Alpert/Christensen (1997), S. 215.

294 Zur Überprüfung der Verteilungsannahme wurde ein Kolmogoroff-Smirnoff-Test durchgeführt. Angenommen werden konnte die Hypothese auf Normalverteilung ($p>0,1$) nur bei der ACA für alle 8 Eigenschaften, bei der CHIC-alt für die Eigenschaften „Abtaumöglichkeit" und „Gefrierfach", bei der CHIC-neu für die Eigenschaft „Gefrierfach" und für die Eigenschaft „Garantiedauer" bei der SE-CCC.

295 Die Anzahl an paarweisen Tests ergibt sich aus $(n(n-1)/2)*8$ Eigenschaften, wobei n für die Anzahl untersuchter Verfahren steht.

Tabelle 5-3: Nicht signifikante Unterschiede zwischen den Verteilungen der Bedeutungsgewichte der verschiedenen Methoden zur Präferenzmessung (Kolmogoroff-Smirnoff-Test p>0,1)

Methode	**ACA**	**CCC-alt**	**CCC-neu**	**CHIC-alt**	**CHIC-neu**	**SE-ACA**	**SE-CCC**	**SE-CHIC**
ACA	——	Preis Inhalt		Energie Garantie Innen				Innen
CCC-alt		——		Marke Preis Energie Inhalt Innen Abtau	Energie		Innen	Marke
CCC-neu			——		Marke Energie			
CHIC-alt				——	Energie		Marke Innen Abtau	Marke Innen Abtau
CHIC-neu					——			
SE-ACA						——	Marke	Marke Inhalt
SE-CCC							——	Marke Preis Energie Garantie Inhalt Innen Abtau Gefrier
SE-CHIC								——

Wie die Ergebnisse in Tabelle 5-3 zeigen, treten nur bei 35 aller 224 getesteten Vergleiche keine signifikanten Unterschiede zwischen den Verteilungen der geschätzten Parameter auf. Berücksichtigt man, daß zwischen der SE-CCC und die SE-CHIC als identische Verfahren die konvergierende Validität getestet wird und folglich keine signifikanten Unterschiede zwischen diesen beiden Methoden zu erwarten sind, kann die Hypothese diskriminierender Validität für 88% der betrachteten Fälle angenommen werden. Eine differenzierte Analyse der Ausnah-

men deutet darauf hin, daß zwischen allen drei Teilstichproben keine signifikanten Abweichungen hinsichtlich der Eigenschaft „Marke“ auftreten. Dieses Ergebnis ist vor allem in Hinblick auf Prüfung der Prognosevalidität positiv zu bewerten, denn damit sinkt die Gefahr, daß die Ergebnisse des Methodenvergleichs durch systematische Unterschiede hinsichtlich der Präferenzen für die betrachteten Marken zwischen den unabhängigen Stichproben verzerrt werden. Weiterhin auffallend ist die relativ hohe Übereinstimmung zwischen den Ergebnissen der CCC-alt und der CHIC-alt. Dies läßt vermuten, daß sich trotz methodischer Unterschiede zwischen den beiden Verfahren vergleichbare Präferenzstrukturen ergeben. Alle übrigen nicht signifikanten Unterschiede weisen auf keine weiteren systematischen Effekte hin, so daß diese als Ausnahmen zu interpretieren sind, für deren Auftreten die Ergebnisse keine plausiblen Begründungen liefern.

Während bisher die untersuchten Eigenschaften jeweils separat betrachtet wurden, soll die diskriminierende Validität in einem weiteren Schritt durch die gleichzeitige Berücksichtigung aller Eigenschaften getestet werden. Hierfür wird der Korrelationskoeffizient nach Pearson zwischen den Mittelwerten der geschätzten Bedeutungsgewichte berechnet.[296] Dieser gibt Auskunft darüber, inwieweit sich die relative Wichtigkeit der 8 Eigenschaften zueinander zwischen den verschiedenen Verfahren unterscheiden. Die Ergebnisse dieser ebenfalls paarweisen Betrachtung sind in Tabelle 5-4 zusammengefaßt.

---

[296] Ein ähnliches Vorgehen zur Überprüfung der Korrespondenz zwischen den ermittelten Bedeutungsgewichten bei unterschiedlichen Verfahren berichten Akaah/Korgaonkar (1983). Die Autoren berechnen die Korrelationen zwischen den Bedeutungsgewichten jedoch zunächst auf individueller Basis und aggregieren die Ergebnisse anschließend. Dies ist allerdings nur möglich, wenn ein within-subject Design bei der Datenerhebung verwendet wurde. In der vorliegenden Untersuchung ist aufgrund der unabhängigen Stichproben zwischen den drei hybriden Verfahren eine entsprechende individuelle Analyse nicht für alle paarweisen Vergleiche durchführbar. Ein weiterer Unterschied zu dem Vorgehen von Akaah/Korgaonkar (1983) liegt darin, daß die Autoren den Rangkorrelationskoeffizienten nach Spearman berechnen (vgl. S. 119). Dieser vergleicht jedoch nur die Reihenfolge der geschätzten Wichtigkeiten. Mit dem Korrelationskoeffizient nach Pearson werden dagegen auch die Unterschiede zwischen den verschiedenen Bedeutungsgewichten berücksichtigt und die Präferenzstrukturen detaillierter analysiert.

Tabelle 5-4: Korrelationskoeffizienten nach Pearson zwischen den Mittelwerten der Bedeutungsgewichte

Methode	ACA	CCC-alt	CCC-neu	CHIC-alt	CHIC-neu	SE-ACA	SE-CCC	SE-CHIC
ACA	——	0,71**	0,57	0,68*	0,62	0,63*	0,45	0,47
CCC-alt		——	0,92***	0,85***	0,87***	0,37	0,70*	0,69*
CCC-neu			——	0,76**	0,89***	0,19	0,61	0,64*
CHIC-alt				——	0,96***	0,66*	0,89***	0,89***
CHIC-neu					——	0,49	0,82**	0,85***
SE-ACA						——	0,75**	0,75**
SE-CCC							——	0,98***
SE-CHIC								——

* $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$

Ein erstes Ergebnis dieser Korrelationsanalyse betrifft die Self-Explicated-Verfahren. So tritt zwischen der SE-CCC und SE-CHIC mit 0,98 eine nahezu perfekte Korrelation zwischen den Bedeutungsgewichten auf und bestätigt dementsprechend die Hypothese konvergierender Validität zwischen identischen Methoden zur Präferenzmessung. Dagegen wirken sich die methodischen Unterschiede der SE-ACA auf die Ergebnisse aus. Ihre Korrelation mit der Self-Explicated-Variante der CCC bzw. CHIC fällt deutlich niedriger aus. Vergleicht man die hybriden Verfahren untereinander, sind die Übereinstimmungen der beiden neuen Verfahren mit den Ergebnissen der ACA sehr ähnlich. Deutlich höher fallen dagegen die Korrelationen zwischen den beiden neuen Methoden aus. Dies gilt sowohl hinsichtlich der beiden Varianten CCC und CHIC als auch hinsichtlich der beiden Schätzverfahren. Letzteres entspricht dabei nicht den Ergebnissen bei separater Analyse der Eigenschaften. Während als Ergebnis der Verteilungstests signifikante Unterschiede zwischen den beiden Schätzverfahren „-alt“ und „-neu“ bei beiden neuen Methoden auftraten (vgl. hierzu Tabelle 5-3), korrelieren die Ergebnisse zwischen CCC-alt und -neu bzw. CHIC-alt und -neu im Gegensatz dazu bei Betrachtung aller Eigenschaften simultan sehr hoch.

Die Überlegungen, die zur Modifikation des Schätzverfahrens führten, wurden in Kapitel 3.1.3 diskutiert. Das Ziel der neuen Berechnungsformel ist, die Teilnutzenwerte von Nicht-Kerneigenschaften nach Kombination der Schätzergebnisse aus beiden Interviewphasen nicht wesentlich stärker einfließen zu lassen als einzelne Kerneigenschaften. Diese Modifikation gewinnt in den Fällen an Bedeutung, in denen einzelne Kerneigenschaften im Conjoint-Teil dominant werden. Während dies in einer vorangegangenen Studie mit Kaffee als Untersuchungsgegenstand vor allem für die Eigenschaften Preis und Marke zutraf, tritt dieser Effekt im vorliegenden Datensatz in der Form nicht auf.[297] Eine mögliche Erklärung hierfür könnte der betrachtete Untersuchungsgegenstand liefern. In der vorliegenden Untersuchung wurde mit Kühlschränken ein Gebrauchsgegenstand betrachtet, bei dem vor allem funktionale Eigenschaften kaufrelevant sind. Im Gegensatz dazu stellt Kaffee ein kurzlebiges Konsumgut dar, bei dem die Schlüsselinformationen Preis und Marke im Verhältnis zu anderen Eigenschaften von hervorragender Bedeutung sind.[298] Dies verdeutlicht, daß die Struktur von Nachfragerpräferenzen auch von der Produktart beeinflußt wird und demzufolge die Bedeutung der vorgeschlagenen Modifikation der Berechnungsformel vom betrachteten Untersuchungsgegenstand abhängen kann.

Als dritter Bereich sei der Zusammenhang zwischen den ermittelten Bedeutungsgewichten der hybriden Verfahren und denen der Self-Explicated-Methoden analysiert. Es zeigt sich, daß die Korrelationen innerhalb der Stichproben für die ACA (0,63*) und CCC (0,7* bzw. 0,61) vergleichbare Werte annehmen. Wesentlich höher und zudem hoch signifikant konvergieren die Bedeutungsgewichte zwischen der SE-CHIC und der CHIC. Eine Erklärung für diese höhere Konvergenz zwischen den beiden Phasen im Gegensatz zur CCC kann darin liegen, daß durch die Verankerung der Choice-Sets zueinander auf Basis der Self-Explicated-Phase mehr Informationen in der Conjoint-Phase integriert werden als dies bei der CCC der Fall ist. Hier beschränkt sich die Verbindung zwischen Self-Explicated-Phase und der Conjoint-Phase auf die Identifikation der Kerneigen-

---

297 Vgl. Hensel-Börner/Sattler (2000), S. 14.

298 Empirische Belege hierfür liefern Bleicker (1983), S. 16; Böcker (1986), S. 543 ff.; Jacoby/Olson/Haddock (1971), S. 570 ff.; Kroeber-Riel (1992), S. 281 ff.; Rao/Sattler (1998) und Sattler (1991), S. 181.

schaften.[299] Des weiteren wird deutlich, daß die Präferenzstrukturen zwischen hybriden Verfahren mit einer anderen Self-Explicated-Methode relativ gering miteinander korrelieren.[300] Dies impliziert, daß Vergleiche zwischen den adaptiven hybriden Verfahren mit einer unabhängigen Self-Explicated-Methode Unterschiede in den ermittelten Präferenzstrukturen aufweisen können.

5.1.3 Zusammenfassung der Ergebnisse hinsichtlich der diskriminierenden Validität

Die diskriminierende Validität wurde für den vorliegenden Datensatz auf zwei verschiedenen Ebenen getestet. Zum einen lieferte das Verfahren CCC die Möglichkeit, den Einfluß der Gestaltung der Beurteilungsaufgabe auf die ermittelten Präferenzstrukturen zu untersuchen. Zum anderen wurde in diesem ersten Schritt der Validitätsprüfung getestet, ob die untersuchten Methoden zur Präferenzmessung zu unterschiedlichen Ergebnissen hinsichtlich der geschätzten Präferenzstrukturen für die befragten Personen führen. Wie bereits erläutert, kann die diskriminierende Validität als Voraussetzung für alle weiteren Validitätsvergleiche angesehen werden. Nur wenn sich Unterschiede in den ermittelten Präferenzstrukturen in Abhängigkeit der eingesetzten Verfahren ergeben, gilt es weiterhin zu testen, welches Verfahren die höchste Übereinstimmung mit den tatsächlichen Präferenzstrukturen der Probanden erreicht.

Zentrale Ergebnisse sind:

1. Zwischen den geschätzten Präferenzstrukturen in Abhängigkeit von der Beurteilungsaufgabe (Rangreihung oder Kaufwahrscheinlichkeitsabfrage) konnte keine diskriminierende Validität nachgewiesen werden. Ein Kolmogoroff-Smirnoff-Test auf Verteilungsunterschiede zeigte nur für 2 der 8 Eigenschaften schwach signifikante Unterschiede, wohingegen bei allen andern Eigenschaften die Hypothese auf Gleichheit in der Verteilung nicht abgelehnt werden konnte. Vor diesem Hintergrund werden bei allen weiteren Analysen lediglich die Ergebnisse der CCC auf Basis der Rangdaten berücksichtigt.

---

299 Vgl. hierzu auch die Argumentation hinsichtlich der Häufigkeiten der Gewichtungsfaktoren w_i* in Kapitel 4.3.3.

300 Einzige Ausnahme bildet hierbei die Konvergenz zwischen der CHIC mit der SE-CCC. Dies ist jedoch plausibel, da die beiden Self-Explicated-Verfahren SE-CCC und SE-CHIC identisch sind und, wie bereits diskutiert, sehr hoch miteinander korrelieren.

2. Die Bedeutungsgewichte auf aggregierter Ebene (Mittelwerte und Standardabweichung) deuten mit Ausnahme der CCC auf relativ homogene Präferenzstrukturen hin. So fallen die durchschnittlichen Bedeutungsgewichte sowohl bei den adaptiven hybriden als auch bei den Self-Explicated-Verfahren für die 8 Eigenschaften vergleichsweise ähnlich aus. Die Standardabweichungen der ermittelten Mittelwerte deuten darauf hin, daß diese Homogenität bei den Self-Explicated-Verfahren wesentlich stärker ausgeprägt ist.
3. Mit Hilfe eines Kolmogoroff-Smirnoff-Tests auf Verteilungsunterschiede konnte bei separater Analyse der einzelnen Eigenschaften gezeigt werden, daß bei individueller Betrachtung in überwiegender Mehrzahl signifikante Unterschiede zwischen den Methoden auftreten. Die einzige Ausnahme zwischen SE-CCC und SE-CHIC bestätigt gleichzeitig, die konvergierende Validität bei identischen Verfahren.
4. Über die Korrelation der aggregierten Mittelwerte der Bedeutungsgewichte wurden bei ebenfalls paarweiser Gegenüberstellung der Verfahren *alle* Eigenschaften simultan berücksichtigt. Mit diesem Kriterium werden die Befunde der individuellen Analyse auf Basis des Verteilungstests nur bedingt bestätigt. Die konvergierende Validität zwischen SE-CCC und SE-CHIC zeigt sich auch hier durch eine fast perfekte Korrelation zwischen den mittleren Bedeutungsgewichten. Abweichend ergibt sich jedoch eine relativ hohe Konvergenz zwischen den alternativen Schätzverfahren bei der CCC und bei der CHIC. Als eine Erklärung wurden die vergleichbaren Bedeutungsgewichte aller Eigenschaften angeführt, die zumindest teilweise auf Besonderheiten des Untersuchungsgegenstandes zurückzuführen sind. Hinsichtlich der Korrelationen zwischen den adaptiven Verfahren und den Self-Explicated-Varianten kann keine eindeutige Aussage getroffen werden. Während die Ergebnisse der ACA und der CCC relativ niedrig mit den Self-Explicated-Varianten korrelieren, weist die CHIC signifikante Korrelationskoeffizienten bei der SE-CHIC und der SE-CCC auf. Als ein Grund hierfür wurde die stärkere Verknüpfung der Informationen aus dem Conjoint-Teil mit denen der Self-Explicated-Phase genannt. Insgesamt ist diesen Teilergebnissen jedoch ein geringeres Gewicht beizumessen, da die individuelle Analyse einen höheren Aussagegehalt liefert.

Die Überprüfung der diskriminierenden Validität ergab, daß keine signifikanten Unterschiede in Abhängigkeit der Präferenzabfrage bei der CCC auftreten. Für

die untersuchten adaptiven Methoden und den jeweiligen Self-Explicated-Verfahren sind auf individueller Ebene signifikante Unterschiede zwischen den geschätzten Präferenzstrukturen festzustellen. Anhand weiterer Gütekriterien muß deshalb überprüft werden, ob und inwieweit sich die Unterschiede auf die Prognosen, basierend auf den geschätzten Präferenzstrukturen, auswirken. Gleichzeitig wird damit untersucht, mit welcher der untersuchten Methoden die tatsächlichen Präferenzstrukturen der Probanden am besten erfaßt werden können.

5.2 Interne Validität

Wie in Kapitel 2 ausgeführt, wird mit der internen Validität die Korrespondenz zwischen den empirisch erhobenen Präferenzdaten und den auf Basis der geschätzten Nutzenwerte ermittelten Präferenzdaten getestet. Überprüft wird, ob und wie gut sich die Inputdaten (empirisch beobachtet) durch die geschätzten Outputdaten rechentechnisch rekonstruieren lassen.[301] Die interne Validität ermöglicht damit eine Aussage über die Güte des Modells an sich. Für die im Rahmen dieser Arbeit untersuchten Verfahren kann diese Überprüfung auf zwei Stufen durchgeführt werden. Sowohl für die CCC als auch für die CHIC kann zunächst die interne Validität im Hinblick auf die LINMAP bzw. OLS-Schätzung innerhalb der Conjoint-Phase ermittelt werden.[302] Anschließend bezieht sich die Analyse für alle drei adaptiven Verfahren auf die Übereinstimmung zwischen beobachteten und geschätzten Werten der abhängigen Variable für das Kalibrierungsset. Die Ergebnisse dieser zweistufigen Überprüfung der internen Validität werden im folgenden dokumentiert.

Das Gütemaß für die LINMAP-Schätzung der CCC bildet der Korrelationskoeffizient nach Kendall. Er wird berechnet zwischen der empirisch beobachteten Rangfolge in der Conjoint-Aufgabe und der auf Basis geschätzter Nutzenwerte errechneten Rangfolge für die Produktkonzepte. Nach dem gleichen Prinzip wird die interne Validität für die OLS-Schätzung der Conjoint-Phase der CHIC gemessen. Hier wird jedoch der Korrelationskoeffizient nach Pearson herangezogen, da die abhängige Variable der geschätzten Regression metrisches Skalenni-

---

301 Vgl. Baier/Säuberlich (1997), S. 968.

302 Für die ACA sieht das System selbst kein vergleichbares Gütekriterium für die Schätzung innerhalb der Paarvergleichsaufgabe vor und kann auch nicht auf Basis des Interviewprotokolls im nachhinein ermittelt werden.

veau aufweist. Für beide Verfahren ergeben sich vergleichbare Wertebereiche hinsichtlich ihrer Gütekriterien. So ergibt sich für die individuellen Kendalls τ–Werte der LINMAP-Schätzung ein Mittelwert von 0,77 (SA=0,21). Für die OLS-Schätzung der Conjoint-Phase der CHIC liegt der durchschnittliche Korrelationskoeffizient nach Pearson für die Stichprobe bei 0,79 (SA=0,13). Eine direkte Gegenüberstellung dieser beiden Gütekriterien für die interne Validität ist jedoch nur bedingt zulässig. Zum einen unterscheiden sich die beiden Parameter hinsichtlich des Skalenniveaus der Inputdaten. Zum anderen hängen beide von der Anzahl der Beobachtungen ab. Während bei der CCC für alle Probanden 9 Beobachtungen in die Schätzung eingingen, variiert die Anzahl der CHIC in Abhängigkeit der individuellen Kerneigenschaften zwischen 8 und 12 Beobachtungen.

Wie bereits oben erwähnt, kann die interne Validität für die adaptiven hybriden Verfahren CCC, CHIC und ACA anhand des Kalibrierungssets getestet werden. Ermittelt wird hierfür die Übereinstimmung zwischen beobachteter Bewertung des Kalibrierungssets und deren Schätzungen auf Basis der kombinierten Teilnutzenwerte. Die Vergleichbarkeit zwischen den drei Verfahren ist jedoch aus den bereits diskutierten Gründen eingeschränkt, denn neben dem Skalenniveau der Präferenzdaten variiert insbesondere die Größe des Kalibrierungssets und damit verbunden die Anzahl der Beobachtungen zwischen 4 (CHIC), 6 (ACA) und 9 (CCC). Die Mittelwerte der jeweiligen Gütemaße sind in Tabelle 5-5 ausgewiesen. Darüber hinaus werden bei der CCC und CHIC die Werte angeführt, die sich bei separater Betrachtung der Self-Explicated-Phase und der Conjoint-Phase ergeben. Entsprechend der in Kapitel 2.2.2.1 vorgenommenen Abgrenzung der verschiedenen Gütekriterien, handelt es sich in diesem Fall nicht um eine Betrachtung der internen Validität im engeren Sinn. Vielmehr kann damit überprüft werden, ob und inwieweit die Kombination der beiden Phasen zu einer höheren Übereinstimmung zwischen geschätzten und prognostizierten Werten für das Kalibrierungsset führt.[303]

---

[303] Auch hier weist die ACA die entsprechenden Informationen und Werte nicht aus, so daß ein Vergleich nicht vorgenommen werden kann.

Tabelle 5-5: Mittelwerte (MW) und Standardabweichungen (SA) der Korrelationskoeffizienten zwischen prognostizierten und beobachteten Werten des Kalibrierungssets

CCC		CHIC		ACA	
Kendall MW (SA)		**Pearson MW (SA)**		**Pearson MW (SA)**	
SE-CCC	0,49 (0,24)	**SE-CHIC**	0,64 (0,49)	**ACA**	0,60 (0,26)
CA-CCC	0,57 (0,28)	**CA-CHIC**	0,48 (0,52)		
CCC	0,68 (0,25)	**CHIC**	0,82 (0,39)		

Tabelle 5-5 zeigt, daß bei beiden neuen Methoden die Kombination der Self-Explicated- und der Conjoint-Phase zu einer höheren Übereinstimmung führt als die jeweils separate Auswertung der beiden Phasen. Dieses Ergebnis war zu erwarten, da das interne Validitätsmaß als Zielkriterium für die Kombination der beiden Phasen herangezogen wurde. Interessant jedoch ist, daß, obwohl die Validierung in einer wiederholten Conjoint-Aufgabe vorgenommen wird, bei der CHIC die Prognose durch die Schätzwerte der Self-Explicated-Phase besser gelingt als durch die Conjoint-Phase. Darüber hinaus ist die Überlegenheit der Self-Explicated-Phase gegenüber der Conjoint-Phase mit einem Anstieg von 0,64 (SE-CHIC) gegenüber 0,48 (CA-CHIC) doppelt so groß als im umgekehrten Fall bei der CCC. Hier kann zwar für die Conjoint-Werte eine höhere Übereinstimmung zwischen prognostizierten und beobachteten Werten festgestellt werden, allerdings fällt dieser Unterschied zwischen 0,57 (CA-CCC) und 0,49 (SE-CCC) deutlich geringer aus als im umgekehrten Fall bei der CHIC. Ein Kolmogoroff-Smirnoff-Test ergab, daß die Unterschiede zwischen den Ergebnissen der einzelnen Phasen hoch signifikant ($p<0{,}01$) sind.[304] In der Tendenz zeigte sich dieses Ergebnis bereits bei der Analyse der optimalen Gewichtungsfaktoren w_i^* in Kapitel 4.3.3. Bei der CCC trat vergleichsweise häufig ein Gewichtungsfaktor nahe 1 auf, wodurch die Conjoint-Teilnutzen die entscheidende Größe für die kombinierten Teilnutzenwerte bilden. Dagegen traf genau der umgekehrte Fall bei der CHIC zu. Dort lag bei fast 20% der Probanden der optimale Gewichtungsfaktor

[304] Auf einen T-Test auf Mittelwertunterschiede wurde verzichtet, da bei allen die Anwendungsvoraussetzung der Normalverteilung nicht gegeben war.

nahe 0.[305] Die Ergebnisse in Tabelle 5-5 zeigen weiterhin, daß die interne Validität insgesamt bei der CHIC mit 0,82 am höchsten ausfällt. Die geringste Validität weist die ACA mit einer durchschnittlichen Korrelation von 0,6 zwischen beobachteten und prognostizierten Werten für das Kalibrierungsset auf, während die CCC mit einem Mittelwert von 0,68 zwischen den beiden anderen Verfahren liegt. Wie bereits erläutert, können diese Werte in ihrer absoluten Höhe nur bedingt direkt miteinander verglichen werden. Ein Test auf Signifikanz der Unterschiede zwischen den adaptiven Verfahren ist aus diesem Grund nicht sinnvoll. Die Überprüfung der internen Validität dient an dieser Stelle vielmehr einer Prüfung der Modellgüte an sich und weniger einem Vergleich zwischen den verschiedenen Verfahren.

Vor diesem Hintergrund muß diskutiert werden, daß die interne Validität sowohl bei den Auswertungen der Conjoint-Phase mit LINMAP und OLS als auch bei allen adaptiven Verfahren relativ niedrig ausfällt. Um sicherzustellen, daß infolgedessen die Gesamtergebnisse insbesondere durch den Einfluß von Ausreißern nicht verzerrt werden, wurde eine entsprechende Sensitivitätsanalyse durchgeführt. Zu diesem Zweck wurden alle drei Teilstichproben zunächst um die Probanden reduziert, bei denen das individuelle Gütemaß für die interne Validität kleiner als 0,2 und in einem zweiten Schritt kleiner als 0,4 lag. Die Werte beziehen sich sowohl auf die interne Validität der Schätzergebnisse innerhalb der Conjoint-Phase bei der CCC und CHIC als auch auf die drei adaptiven Verfahren, gemessen an der Prognosegüte des Kalibrierungssets. Die jeweils verbleibende Anzahl an Probanden in den Teilstichproben ist in Tabelle 5-6 wiedergegeben.

---

[305] Vgl. hierzu auch Tabelle 4-4, in der eine Häufigkeitsverteilung über die optimalen Gewichtungsfaktoren w_i^* für alle Probanden bei der CCC und der CHIC angegeben wird.

Tabelle 5-6: Umfang der Teilstichproben nach Eliminierung von Ausreißern

Stichprobenumfang \ Methode		CCC	CHIC	ACA
Gesamt-stichprobe	N	242	234	249
Stichprobe bei einem Grenzwert für das interne Gütemaß bei 0,2	N	225	217	220
Stichprobe bei einem Grenzwert für das interne Gütemaß bei 0,4	N	205	213	178

Um die Stabilität der Gesamtergebnisse hinsichtlich der Probanden mit sehr geringer interner Validität zu testen, diente ein Kolmogoroff-Smirnoff-Test auf Verteilungsunterschiede zwischen den Bedeutungsgewichten für jede Eigenschaft. Auch hier wurden die geschätzten Präferenzstrukturen betrachtet, da sie die Grundlage für alle weiteren Analysen und Validitätsvergleiche bilden.[306] Bei jedem Verfahren wurde die Gesamtstichprobe (max) jeweils mit den um die Ausreißer reduzierten Stichproben (0,2 und 0,4) verglichen. Für alle durchgeführten Verteilungstests ergab sich, daß nach Eliminierung von Probanden mit niedriger interner Validität keine signifikanten Unterschiede ($p>0,1$) zwischen der Gesamtstichprobe und den jeweils reduzierten auftreten. Dies trifft bei allen Methoden hinsichtlich aller 8 Eigenschaften gleichermaßen zu. Die mangelnde Vergleichbarkeit der internen Gütemaße bedeutet allerdings gleichzeitig, daß die gewählten Grenzwerte für Ausreißer nicht problemlos auf alle Verfahren übertragbar sind. Eine Eliminierung von Probanden mit niedriger interner Validität ist demzufolge nur über ein hohes Maß an Subjektivität zu realisieren. Um eine dadurch verursachte Verzerrung zu vermeiden und da es auf das Gesamtergebnis, wie gesehen, keinen Einfluß hat, werden alle weiteren Analysen auf Basis der jeweiligen Gesamtstichprobe durchgeführt. Weiterhin wird durch die Berücksichtigung aller Probanden sichergestellt, daß die Vergleichbarkeit der Stichpro-

[306] Vgl. hierzu auch die Argumentation in Zusammenhang mit der Überprüfung der diskriminierenden Validität in Kapitel 5.1.

ben in ihrer demografischen Struktur erhalten bleibt. Auch das ist eine notwendige Voraussetzung für die Gültigkeit des Verfahrensvergleichs.

5.3 Prognosevalidität

Ergebnis der diskriminierenden Validitätsprüfung war, daß verschiedene Methoden zur Präferenzmessung zu unterschiedlichen Schätzergebnissen führen. Eine Aussage darüber, welche der untersuchten Methoden die tatsächlichen Präferenzstrukturen der Probanden am besten abbildet, kann jedoch noch nicht gemacht werden. Gleiches gilt für die Befunde der internen Validität, da die dort eingesetzten Gütemaße nur bedingt über die verschiedenen Methoden vergleichbar sind (vgl. hierzu Kapitel 5.2). Um jedoch eine entsprechende Aussage zu ermöglichen, werden die Verfahren im folgenden hinsichtlich ihrer Prognosegüte verglichen. Entscheidendes Kennzeichen dieses Gütekriteriums ist, daß Präferenzen prognostiziert werden, die in keiner Phase der Datenauswertung in die Schätzung der Teilnutzenwerte eingehen. Üblicherweise werden den Probanden hierfür zusätzliche Produktkonzepte zur Beurteilung vorgelegt und zusätzliche Präferenzen der Probanden erhoben. Für die Validierung wird dabei angenommen, daß diese die tatsächlichen Präferenzen der Probanden zeigen und die untersuchten Methoden danach beurteilt, wie gut diese Präferenzen vorhergesagt werden können.[307] Die zusätzlich zu bewertenden Produktkonzepte werden auch Holdout-Konzepte bzw. als Holdout-Set bezeichnet. Zwar wird die Validierung der Ergebnisse anhand von Holdout-Daten in der Literatur nicht eindeutig einem Gütemaß zugeordnet,[308] dennoch basiert der größte Teil bisheriger Validitätsvergleiche auf dieser Vorgehensweise.[309]

Zur Quantifizierung der Prognosevalidität werden nachstehend drei alternative Operationalisierungen angewendet, deren Ergebnisse im folgenden behandelt werden.

---

[307] Vgl. Johnson (1989), S. 273.

[308] Vgl. hierzu Kapitel 2.2.2.1.

[309] Vgl. Green/Srinivasan (1990), S. 13; Wittink/Bergestuen (1999), S. 10; Srinivasan/deMaCarty (1998), S. 2 sowie die Dominanz dieses Kriteriums im Literaturüberblick der vorliegenden Arbeit.

5.3.1 First-Hit-Rate

Wie in Kapitel 4.3.4 beschrieben, bestand die Beurteilungsaufgabe der Probanden hinsichtlich der Holdout-Sets darin, die Rangfolge anzugeben, in der sie die 5 Kühlschränke am ehesten kaufen würden. Um die Holdout-Aufgabe als simulierte Kaufentscheidung interpretieren zu können, wird zunächst lediglich die Information hinsichtlich des ersten Rangplatzes herangezogen. Dabei wird angenommen, daß die Probanden bei einer Kaufentscheidung das Produktkonzept auswählen würden, dem sie den ersten Rang zugeordnet haben. Diese simulierte Kaufentscheidung läßt sich als Maß für die Prognosevalidität der untersuchten Verfahren verwenden. Hierzu werden auf Basis der geschätzten Teilnutzenwerte für die 5 realen Angebote Gesamtnutzenwerte berechnet.[310] Mit der First-Hit-Rate wird für jede Methode getestet, ob dem Produkt mit dem höchsten prognostizierten Gesamtnutzen auch tatsächlich der erste Rang zugeordnet wurde. In Tabelle 5-7 sind die entsprechenden Prozentsätze für die richtig prognostizierten Rangplätze ausgewiesen. Für alle acht untersuchten Methoden zeigt sich eine deutliche Abweichung von der Idealanpassung (100%). Vorangegangene Untersuchungen betonen jedoch, daß eine solche 100%ige Anpassung für die Prognose des ersten Rangplatzes nicht zu erwarten ist.[311] Dennoch zeigt sich, daß die ermittelten Prozentwerte der First-Hit-Rate z.T. deutlich niedriger ausfallen als bei den angeführten bisherigen Untersuchungen.

---

[310] Da die Marke „Electrolux" nicht als Ausprägung in den durchgeführten Conjoint-Interviews enthalten war, wurden Informationen aus dem abschließenden Interviewteil zur Markeneinschätzung herangezogen. Hierbei wurden die Probanden gebeten, zusätzliche Marken, die nicht als Ausprägung im eigentlichen Befragungsprogramm enthalten waren, auf einer Rating-Skala 0 bis 10 zu bewerten. Diese Vorgehensweise entsprach der ersten Phase der Self-Explicated-Aufgabe der hybriden Verfahren. Somit konnten die Teilnutzenwerte für die Ausprägung „Electrolux" nachträglich je nach vorgenommener Bewertung durch die Probanden inter- oder extrapoliert werden.

[311] Vgl. z.B. Goldberg (1980): [22%-32%], zitiert nach Green/Goldberg/Montemayer (1981), S. 39; Albers/Brockhoff (1985): [60%], S. 197; Mehta/Moore/Pavia (1992): [„...models predicted approximately half of the first choices correctly"], S. 473; Huber et al. (1993): [56%-70%], S. 110; Sattler (1994): [76%-79%], S. 37; Srinivasan/Park (1997): [64%-76%], S. 289; Hensel-Börner/Sattler (2000): [42%-57%] S. 16.

Tabelle 5-7: Ergebnisse der Prognosevalidität, gemessen durch die First-Hit-Rate

Methode	**ACA**	**CCC-alt**	**CCC-neu**	**CHIC-alt**	**CHIC-neu**	**SE-ACA**	**SE-CCC**	**SE-CHIC**
N	249	242	242	234	234	249	242	234
First-Hit-Rate a)	45,4%	44,2%	41,3%	39,3%	44,9%	40,2%	39,7%	41,5%
Verbesserung gegenüber dem Zufallsmodell	31,8% ***	30,3% ***	26,6% ***	24,1% ***	31,1% ***	25,3% ***	24,6% ***	26,9% ***

a) Angegeben wird der Anteil an Personen, bei denen für das auf Rang 1 bewertete Produkt der erste Rang richtig prognostiziert wurde.
*** $p<0,01$

Die insgesamt beste Anpassung wird mit 45,4% bei der ACA erzielt. Die CCC-alt und CHIC-neu erreichen mit 44,2% bzw. 44,9% vergleichbare Werte. Alle anderen Verfahren schwanken mit 39,3% (CHIC-alt) und 41,5% (SE-CHIC) in einem relativ kleinen Intervall und weisen somit eine niedrigere Anpassung als die drei genannten hybriden Verfahren auf. Überraschend hierbei ist, daß die modifizierte Schätzmethode bei der CHIC die erwartete Verbesserung in der Prognosegüte bewirkt, bei der CCC dagegen die First-Hit-Rate für die neue Variante geringer ausfällt. Ein zusätzlicher Vergleich der Methoden gegenüber einem Zufallsmodell[312] ergab, daß die Verbesserungen bei allen Methoden hoch signifikant ($p<0,01$) sind.

Auch wenn ein direkter Vergleich der vorliegenden Ergebnisse mit den Befunden bisheriger Studien nicht unproblematisch ist, können die z.T. deutlichen Unterschiede hinsichtlich der First-Hit-Rate nur bedingt durch Besonderheiten im Untersuchungsdesign erklärt werden. Ein Grund für die relativ niedrigen Anpassungen kann vielmehr in mangelnder Reliabilität der Antworten liegen.[313] Sofern Probanden bei einer Wiederholung der Beurteilungsaufgabe anders geantwortet haben als im eigentlich Conjoint-Interview, können dadurch deutliche Abwei-

---

312 Zufallsmodell: Wahrscheinlichkeit, daß die Prognose für den ersten Rang mit der tatsächlichen Entscheidung übereinstimmt. Für die vorliegende Untersuchung beträgt diese Wahrscheinlichkeit 20%, da im Holdout-Set insgesamt 5 Produkte zu bewerten waren. Prozentuale Verbesserung = 100*[(korrekte Prognose durch die Präferenzmeßmethode in Prozent – korrekte Prognose durch das Zufallsmodell in Prozent) / (100 – korrekte Prognose durch das Zufallsmodell in Prozent)]. Vgl. Srinivasan/Park (1997), S. 289.

chungen in den Prognoseergebnissen auftreten. Eine mögliche Ursache für solch inkonsistentes Antwortverhalten kann die Positionierung der Holdout-Aufgabe im Interviewablauf sein. Die erneute Abfrage der Präferenzen am Ende des Gesamtinterviews birgt die Gefahr, die Auskunftsbereitschaft der Probanden zu überfordern und die Gültigkeit der gegebenen Antworten dadurch einzuschränken. Weiterhin kann nicht ausgeschlossen werden, daß die Bewertungen für das Holdout-Set davon beeinflußt wurden, daß die Produktalternativen nur durch 5 der untersuchten 8 Eigenschaften beschrieben waren. Insbesondere bei Probanden, bei denen die nicht berücksichtigten Eigenschaften maßgeblich für die Kaufentscheidung sind, kann dies die Gefahr willkürlicher Antworten vergrößert haben.

Die beobachteten Unterschiede in der First-Hit-Rate zwischen den Methoden wurden durch einen Vergleich der Parameter zweier Binomialverteilungen zusätzlich auf Signifikanz getestet.[314] Die Irrtumswahrscheinlichkeiten für die Testgröße dieses paarweise durchgeführten χ^2-Tests sind in Tabelle 5-8 angegeben.

Tabelle 5-8: Irrtumswahrscheinlichkeiten p der χ^2-Tests auf Signifikanz der Unterschiede zwischen der First-Hit-Rate

Methode	**ACA**	**CCC-alt**	**CCC-neu**	**CHIC-alt**	**CHIC-neu**	**SE-ACA**	**SE-CCC**	**SE-CHIC**
ACA	——	p=0,79	p=0,36	p=0,18	p=0,91	p=0,24	p=0,20	p=0,38
CCC-alt		——	p=0,52	p=0,28	p=0,89	p=0,36	p=0,31	p=0,54
CCC-neu			——	p=0,79	p=0,43	p=0,79	p=0,71	p=0,98
CHIC-alt				——	p=0,22	p=0,85	p=0,94	p=0,64
CHIC-neu					——	p=0,30	p=0,25	p=0,46
SE-ACA						——	p=0,91	p=0,77
SE-CCC							——	p=0,69
SE-CHIC								——

[313] Vgl. Johnson (1989), S. 273.

Die paarweisen χ^2-Tests belegen, daß alle aufgetretenen Unterschiede zwischen den Verfahren in der First-Hit-Rate bei einer Irrtumswahrscheinlichkeit von $p>0{,}10$ nicht signifikant sind. Insgesamt kann den hybriden Verfahren (mit Ausnahme der CHIC-alt) eine leicht höhere Prognosevalidität, gemessen durch die First-Hit-Rate, als den Self-Explicated-Verfahren bescheinigt werden. Es zeigt sich aber, daß diese Überlegenheit relativ gering ausfällt und zudem nicht signifikant ist. Dieses Ergebnis entspricht in seiner Tendenz den von Srinivasan/Park (1997) berichteten Befunden. Die Autoren können bei vergleichbarer Vorgehensweise zur Operationalisierung der Prognosegüte keine signifikanten Unterschiede zwischen der von ihnen vorgeschlagenen CCA und den Self-Explicated-Modellen nachweisen.[315]

5.3.2 Individuelle Kendalls τ-Werte

Wie bereits erläutert, bestand die Beurteilungsaufgabe für die Probanden hinsichtlich des Holdout-Sets in der Rangreihung der 5 Produktalternativen, so daß die Prognosevalidität in einem nächsten Schritt auf Basis dieser zusätzlichen Informationen überprüft werden kann. Hierfür wurde über die ermittelten Gesamtnutzenwerte eine individuelle Präferenzrangfolge für jede Testperson bzgl. der 5 Produkte abgeleitet, um diese mit der empirisch beobachteten Rangfolge vergleichen zu können. Als Gütekriterium für die Übereinstimmung wurde der individuelle Korrelationskoeffizient nach Kendall zwischen beiden Rangfolgen berechnet und für jedes Verfahren der Mittelwert über die entsprechende Teilstichprobe gebildet (vgl. Tabelle 5-9).

---

[314] Zur Vorgehensweise dieses Tests vgl. Hartung (1991), 496 ff.

[315] Vgl. Srinivasan/Park (1997), S. 290. In ihrer Untersuchung ergibt sich für die Self-Explicated-Methode sogar eine leicht höhere First-Hit-Rate als für die CCA. Es sei zudem angemerkt, daß die Autoren den Vergleich nicht anhand eines zusätzlichen Holdout-Sets, sondern auf Basis tatsächlich getroffener Entscheidungen der Probanden durchführen.

Tabelle 5-9: Mittelwerte und Standardabweichungen der ermittelten Korrelationskoeffizienten (Kendalls τ) zwischen beobachteter und prognostizierter Rangfolge des Holdout-Sets

Methode	N	Kendalls τ	
		Mittelwerte	Standardabweichungen
ACA	249	0,32	0,42
CCC-alt	242	0,33	0,40
CCC-neu	242	0,33	0,41
CHIC-alt	234	0,26	0,43
CHIC-neu	234	0,27	0,41
SE-ACA	249	0,28	0,41
SE-CCC	242	0,28	0,39
SE-CHIC	234	0,29	0,40

Auffallend bei allen untersuchten Methoden ist, daß die durchschnittlichen Kendalls τ-Werte sehr niedrig sind. Dies impliziert, daß die Prognose der Rangfolge für das Holdout-Set auf Basis der geschätzten Teilnutzenwerte im Mittel nur unzureichend gelingt. Die hohen Standardabweichungen deuten zudem darauf hin, daß die Ergebnisse bei allen Methoden sehr stark um diese Mittelwerte streuen. Auch wenn andere Autoren vergleichbar niedrige Werte hinsichtlich dieses Gütekriteriums nennen,[316] stützen die Ergebnisse die Vermutung, daß Reliabilitätsprobleme im Rahmen der Holdout-Aufgabe aufgetreten sind. Wie bereits im Zusammenhang mit der First-Hit-Rate erläutert, können dadurch die Prognoseergebnisse der einzelnen Verfahren beeinträchtigt werden. Eine weitere mögliche Erklärung für die extrem niedrigen Korrelationen zwischen prognostizierten und beobachteten Rangfolgen für das Holdout-Set kann darin liegen, daß die schriftliche Erfassung der Antworten, im Gegensatz zur PC-gestützten bei allen anderen Interviewphasen, die Ergebnisse beeinflußt hat. Im schlechtesten Fall können organisatorische Fehler bei der Datenerhebung z.B. in der Zuordnung der schrift-

[316] Vgl. Goldberg (1980), zitiert nach Green/Goldberg/Montemayer (1981), S. 39; Akaah/Korgaonkar (1983), S. 192 und Green/Shaffer/Patterson (1991), S. 313. Einschränkend muß jedoch angemerkt werden, daß nur in der Studie von Akaah/Korgaonkar (1983) mit 6 Stimuli ein ähnlich großes Holdout-Set prognostiziert wurde. In den beiden anderen Untersuchungen umfaßte die Holdout-Aufgabe mit 9 bzw. 12 deutlich mehr Stimuli, wodurch ein direkter Vergleich der Ergebnisse nur bedingt zulässig ist.

lichen Fragebögen zu den PC-Interviews aufgetreten sein. So vermuten beispielsweise auch Akaah/Korgaonkar (1983) aufgrund der sehr niedrigen Prognosegüte für das Holdout-Set in ihrer Studie „some noise in the data".[317]

Trotz dieser sehr niedrigen Prognosevalidität in ihrer absoluten Höhe, gemessen durch die individuellen Kendalls τ–Werte, läßt sich als weiteres Ergebnis festhalten, daß für alle 8 untersuchten Methoden die Mittelwerte über die Gesamtstichproben nahezu gleich ausfallen. Dies bestätigen auch die Kolmogoroff-Smirnoff-Tests auf Verteilungsunterschiede der individuellen Kendalls τ–Werte zwischen den Methodenpaaren. In Tabelle 5-10 sind die kritischen Irrtumswahrscheinlichkeiten p für die jeweilige Teststatistik ausgewiesen.

Tabelle 5-10: Ergebnisse der Kolmogoroff-Smirnoff-Tests auf Unterschiede in den Verteilungen der individuellen Kendalls τ-Werte für das Holdout-Set

Methode	**ACA**	**CCC-alt**	**CCC-neu**	**CHIC-alt**	**CHIC-neu**	**SE-ACA**	**SE-CCC**	**SE-CHIC**
ACA	——	p=0,99	p=1,00	p=0,14	p=0,23	p=0,76	p=0,11	p=0,93
CCC-alt		——	p=1,00	p=0,15	p=0,52	p=0,73	p=0,32	p=0,84
CCC-neu			——	p=0,22	p=0,48	p=0,87	p=0,27	p=0,89
CHIC-alt				——	p=1,00	p=0,69	p=0,86	p=0,86
CHIC-neu					——	p=1,00	p=0,97	p=0,86
SE-ACA						——	p=0,94	p=1,00
SE-CCC							——	p=0,61
SE-CHIC								——

Die Ergebnisse in Tabelle 5-10 unterstreichen, daß die Nullhypothese gleicher Verteilungen bei allen Paarvergleichen nicht abgelehnt werden kann ($p>0,1$) und folglich keine signifikanten Unterschiede zwischen den Methoden auftreten. Dieses zweite Ergebnis ist vor allem für den Vergleich der Methoden von zentraler

[317] Akaah/Korgaonkar (1983), S. 195.

Bedeutung. Es läßt vermuten, daß im Falle nicht beobachtbarer Einflüsse, z.B. aufgrund der Durchführung der Datenerhebung, das relative Abschneiden der verschiedenen Methoden nicht systematisch verzerrt wurde. Sind Reliabilitätsprobleme hauptsächlich für die niedrige Prognosevalidität verantwortlich, kann angenommen werden, daß diese zumindest im Mittel bei allen drei Stichproben in vergleichbarem Umfang aufgetreten sind. Andernfalls wären wesentlich größere Unterschiede zwischen den verschiedenen Methoden zu beobachten. Vor dem Hintergrund dieses zweiten Ergebnisses kann davon ausgegangen werden, daß die Aussagen hinsichtlich des Verfahrensvergleichs trotz der niedrigen Prognosegüte aller Methoden zumindest in ihrer Tendenz bestehen bleiben können.

Die bisherigen Ergebnisse der individuellen[318] Gütekriterien zur Prognosevalidität deuten darauf hin, daß sich die in Kapitel 5.1 ebenfalls auf individueller Ebene nachgewiesenen signifikanten Unterschiede zwischen den ermittelten Präferenzstrukturen nicht auf die Prognosegüte der untersuchten Verfahren auswirkt. Vielmehr kann bei individueller Auswertung der Daten des Holdout-Sets geschlußfolgert werden, daß die verschiedenen Verfahren zu vergleichbaren Ergebnissen hinsichtlich der Prognosevalidität führen.

5.3.3 Simulierte Marktanteile

Die Bewertung des Holdout-Sets durch die Probanden ermöglicht neben einer individuellen Analyse auch die Bestimmung simulierter Marktanteile der betrachteten Marken. Die beobachteten Marktanteile ergeben sich aus der Anzahl an Probanden, die einer Marke den ersten Rangplatz zugeordnet haben. Wird auf Basis der geschätzten Teilnutzenwerte zudem eine Prognose der Marktanteile vorgenommen, liefert ein Vergleich dieser mit den empirisch beobachteten eine weitere Möglichkeit, jetzt auf aggregierter Ebene, zur Überprüfung der Prognosevalidität. Voraussetzung für die Prognose der entsprechenden Marktanteile ist

---

318 Auch wenn bei der First-Hit-Rate der Anteil der Personen betrachtet wird, bei denen die Prognose übereinstimmt, gehen dabei die Informationen jedes einzelnen Probanden ein. Für jeden wird individuell überprüft, ob die Prognose für den ersten Rangplatz mit dessen tatsächlicher Bewertung übereinstimmt. Im Gegensatz dazu wird bei der folgenden Betrachtung simulierter Marktanteile nicht mehr jeder Proband einzeln analysiert, sondern die Teilstichprobe als ganzes betrachtet. Insofern kann die First-Hit-Rate als individuelles Gütekriterium für die Prognosevalidität interpretiert werden (vgl. hierzu auch Huber et al. (1993), die von „...individual hit rates..." sprechen (S. 109) und Wittink/Bergestuen (1999), S. 11).

eine Transformation der geschätzten Gesamtnutzenwerte in Kaufwahrscheinlichkeiten. In der Literatur werden diesbezüglich alternative Kaufverhaltensmodelle vorgeschlagen, die im folgenden kurz erläutert werden.

5.3.3.1 Kaufverhaltensannahmen[319]

Auf Basis eines Kaufverhaltensmodells wird festgelegt, wie die geschätzten Gesamtnutzenwerte in Kaufwahrscheinlichkeiten transformiert werden. Eine solche Transformation ist erforderlich, da der geschätzte Gesamtnutzen ein Konstrukt darstellt, um die Präferenzen der Probanden zu quantifizieren. Er läßt in der Form noch keine Aussage hinsichtlich der Kaufwahrscheinlichkeit für eine Produktalternative zu. Eine der einfachsten Kaufverhaltensannahmen ist das sog. First-Choice-Model,[320] welches implizit bereits bei der Überprüfung der First-Hit-Rate verwendet wurde:

- First-Choice-Rule oder Maximale-Nutzenregel:[321]

$$(5\text{-}1)\quad p_{im} = \begin{cases} 1; \text{ wenn } U_{im} = MAX\{U_{il} | l \in M\} \\ 0; \text{ sonst} \end{cases} \qquad (i \in I;\ m \in M)$$

wobei:

p_{im}: Kaufwahrscheinlichkeit für die m-te Alternative für den i-ten Probanden,

U_{im}: Geschätzter Gesamtnutzen für die m-te Alternative für den i-ten Probanden,

I: Indexmenge der Probanden,

M: Indexmenge der Alternativen.

Nach dem First-Choice-Model wird die Produktalternative mit einer Wahrscheinlichkeit von 100% gewählt, die unter den zur Verfügung stehenden Alternativen den höchsten Gesamtnutzen aufweist. Alle anderen Alternativen erhalten eine Kaufwahrscheinlichkeit von 0%. Die Vorteile des First-Choice-Model liegen vor allem in seiner konzeptionellen wie auch schätztechnischen Einfachheit.

---

319 Einen sehr guten Überblick über die verschiedenen Kaufverhaltensannahmen geben Green/Krieger (1988).

320 Vgl. Finkbeiner (1988), S. 78.

321 Vgl. Green/Krieger (1988), S. 116.

Gleichzeitig neigt diese Transformationsvorschrift allerdings dazu, die Kaufwahrscheinlichkeiten für attraktivere Produkte zu überschätzen, wohingegen für weniger attraktive die Kaufwahrscheinlichkeit unterschätzt wird.[322] Dieser Nachteil gewinnt insbesondere dann an Bedeutung, wenn die Unterschiede zwischen den geschätzten Gesamtnutzen relativ gering sind. In diesem Fall erhält ein Produkt mit nur geringem Mehrnutzen gegenüber den anderen Alternativen eine Kaufwahrscheinlichkeit von 100%.[323] Dieser Punkt macht deutlich, daß das First-Choice-Model voraussetzt, daß die Schätzung der Teilnutzenwerte verzerrungsfrei und ohne Schätzfehler gelingt.[324]

Im Gegensatz dazu werden bei alternativen Kaufverhaltensmodellen die Kaufwahrscheinlichkeiten konkurrierender Produktalternativen entsprechend ihrer geschätzten Gesamtnutzen aufgeteilt. Dabei wird berücksichtigt, daß nicht mit Sicherheit erwartet werden kann, welche der Produktalternativen ausgewählt wird und demzufolge auch Produkte mit geringerem Gesamtnutzen eine Kaufwahrscheinlichkeit größer 0% aufweisen. Ferner wird auch der Tatsache Rechnung getragen, daß Probanden nicht immer das Produkt mit dem höchsten Nutzen auswählen bzw. gar keine eindeutige „erste Wahl" haben.[325] Dies trifft vor allem bei kurzlebigen Konsumgütern zu.[326] Für die Bestimmung der Kaufwahrscheinlichkeiten unter Berücksichtigung der geschätzten Gesamtnutzenwerte liefert die Literatur alternative Modellierungsformen, die sich formal wie folgt darstellen lassen:

- Bradley-Terry-Luce-Kaufwahrscheinlichkeit (BTL-Model)[327]

$$(5\text{-}2) \quad p_{im} = \frac{U_{im}}{\sum_{l \in M} U_{il}} \qquad (i \in I;\ m \in M)$$

---

322 Vgl. Curry (1998), S. 2.

323 Vgl. hierzu auch Gutsche (1995), S. 142 f.

324 Vgl. Elrod/Kumar (1989), S. 260.

325 Vgl. Curry (1998), S. 2. Auch Elrod/Kumar (1989) führen im Zusammenhang mit den Nachteilen des First-Choice-Model als bereits bekanntes Phänomen an, daß die Probanden nicht immer die meist präferierte Alternativen auswählen (vgl. S. 260).

326 Vgl. Green/Krieger (1988), S. 118.

327 Vgl. Green/Krieger (1988), S. 115 f.

- Exponentiell gewichtete Kaufwahrscheinlichkeit[328]

$$(5\text{-}3) \quad p_{im} = \frac{U_{im}^{\beta}}{\sum_{l \in M} U_{il}^{\beta}} \qquad (i \in I;\ m \in M)$$

- Logit-Kaufwahrscheinlichkeit[329]

$$(5\text{-}4) \quad p_{im} = \frac{e^{U_{im}}}{\sum_{l \in M} e^{U_{il}}} \qquad (i \in I;\ m \in M)$$

Diese drei Kaufverhaltensregeln implizieren, daß sich ein höherer Gesamtnutzen eines Produkts in einer entsprechend höheren Kaufwahrscheinlichkeit niederschlägt. Weiterhin wird sichergestellt, daß sich die Kaufwahrscheinlichkeiten über alle betrachteten Alternativen zu 1 addieren.[330] Als nachteilig dagegen ist insbesondere für das BTL- und das Logit-Modell die Besonderheit zu bewerten, daß sich häufig für alle betrachteten Alternativen relativ ähnliche Kaufwahrscheinlichkeiten ergeben. Dadurch steigt bei Hinzunahme zusätzlicher Alternativen mit ähnlichen Gesamtnutzen die Kaufwahrscheinlichkeit für diese überproportional an.[331] Ausgeglichen werden kann dieser Effekt durch das exponentiell gewichtete Modell anhand des Exponenten β, denn je größer der Wert hierfür gewählt wird, desto stärker sind die Unterschiede in den ermittelten Kaufwahrscheinlichkeiten für die konkurrierenden Alternativen ausgeprägt. Im Extremfall, d.h. für sehr große β-Werte, nähert sich das exponentiell gewichtete Modell an das First-Choice-Model an.[332] Wird dagegen der Exponent $\beta=1$ gesetzt, entspricht das exponentiell gewichtete Kaufverhaltensmodell dem BTL-Model.[333]

Auch wenn diese vier Kaufverhaltensannahmen nicht alle in der Literatur diskutierten Modelle repräsentieren, so stellen sie sowohl in der Wissenschaft als auch

---

328 Vgl. Green/Krieger (1988), S. 116.
329 Vgl. auch Hahn (1997), S.156.
330 Vgl. Curry (1998), S. 3.
331 Zu den Problemen der BTL- und der Logit-Regel vgl. Green/Krieger (1988), S. 116.
332 Vgl. Green/Krieger (1988), S. 116.
333 Veranschaulichende Zahlenbeispiele zum Einfluß der Kaufverhaltensannahme auf die resultierenden Marktanteile gibt Curry (1998).

in der kommerziellen Anwendung die am häufigsten eingesetzten Modelle dar.[334] Die Diskussion der jeweiligen Vor- und Nachteile macht aber bereits deutlich, daß nicht a priori bestimmbar ist, welchem Modell der Vorzug gegeben werden kann.[335] Aus diesem Grund wurden zur Prognose der Marktanteile die angeführten Modelle alternativ verwendet. Für das exponentiell gewichtete Modell wurden unterschiedliche Wertebereiche für β angenommen, da keine eindeutige a priori Festlegung vorgenommen werden kann. Vielmehr gilt es, den Exponenten β zu bestimmen, bei dem die höchste Übereinstimmung zwischen prognostizierten und beobachteten Marktanteilen erzielt werden kann. Als Wertebereiche für β werden drei Intervalle betrachtet, die in jeweils kleinere Teilschritte untergliedert sind. Es sind dies Wertebereiche zwischen (0,1; 0,2; ...;1); (2;3; ...;10) und (20; 30; ...; 100), in denen β systematisch variiert.

Die simulierten Marktanteile wurden somit für alle untersuchten Methoden zur Präferenzmessung unter Annahme von insgesamt 30 verschiedenen Kaufverhaltensmodellen prognostiziert.[336] Ausgangspunkt war die Ermittlung der individuellen Kaufwahrscheinlichkeiten für die 5 Alternativen des Holdout-Sets. Nach Aggregation dieser individuellen Werte liegt für jede der untersuchten Methoden eine Prognose der Marktanteile für die im Rahmen der Holdout-Aufgabe betrachteten Marken vor. Wie bereits erwähnt, ergeben sich die empirisch beobachteten Marktanteile aus der Anzahl der Probanden, die einer Marke in der Bewertung der Holdout-Konzepte den ersten Rang zugeordnet haben.

Gemessen wird die Prognosevalidität bei dieser aggregierten Betrachtung anhand von zwei Gütekriterien. Zunächst wurde der Korrelationskoeffizient nach Pearson zwischen den 5 Marktanteilen berechnet. Damit wird jedoch lediglich erfaßt, wie gut die Relationen der unterschiedlichen Marken zueinander prognostiziert werden können. Abweichungen in der absoluten Höhe bleiben unberücksichtigt.[337] Aufgrund dieser für die Prognose von Marktanteilen eingeschränkten Aus-

---

334 Vgl. hierzu auch Green/Krieger (1988), S. 115.

335 Vgl. Green/Krieger (1988), S. 117 ff.

336 Angemerkt sei, daß lediglich vier verschiedene Kaufverhaltensmodelle im engeren Sinn verglichen wurden. Die Anzahl von 30 ergibt sich aus der Tatsache, daß für das exponentiell gewichtete Modell insgesamt 27 verschiedene Werte für den Exponenten β angenommen wurden.

337 Dieser Effekt soll an einem einfachen Zahlenbeispiel veranschaulicht werden. Für die bei-

sagefähigkeit des Korrelationskoeffizienten wurde zusätzlich die mittlere absolute Abweichung (MAA) zwischen den beobachteten und prognostizierten Marktanteilen ermittelt. Dieser Wert gibt Auskunft darüber, wie gut die Marktanteile in ihrer absoluten Höhe prognostiziert werden können. Die Darstellung der Ergebnisse dieser aggregierten Validitätsprüfung ist im folgenden zweigeteilt. Zunächst wird für jede Methode das Kaufverhaltensmodell angenommen, bei dem die höchste Prognosevalidität, gemessen auf Basis von Pearson und der MAA, erzielt wird. In einem zweiten Schritt wird die Stabilität der Ergebnisse in Abhängigkeit der Kaufverhaltensannahme analysiert.

5.3.3.2 Maximale Prognosevalidität der verschiedenen Methoden

In Tabelle 5-11 sind für jedes Verfahren die maximalen Ausprägungen der beiden Gütekriterien Pearson und MAA ausgewiesen.[338] Gleichzeitig wird dabei angegeben, bei welcher Kaufverhaltensannahme sich dieser Maximalwert ergeben hat.

---

den Zahlenreihen (1;2;3;4;5) und (10;20;30;40;50) ergibt sich ein Korrelationskoeffizient nach Pearson von 1, welcher eine perfekte Übereinstimmung impliziert. Keinen Einfluß hat, daß die Werte der zweiten Zahlenreihe um das Zehnfache größer sind als die der ersten.

338 Für die mittlere absolute Abweichung entspricht dieser „Maximalwert" dem Minimum aller ermittelten Werte.

Tabelle 5-11: Maximale Übereinstimmungen zwischen beobachteten und prognostizierten Werten der simulierten Marktanteile gemessen durch Pearson und mittlere absolute Abweichung (MAA)

Methode	Pearson	Kaufverhaltensmodell	MAA in %	Kaufverhaltensmodell
ACA	0,64	exp. gew. Modell: β=0,2	5,9	exp. gew. Modell: β=3
CCC-alt	0,79	exp. gew. Modell: β=0,1	5,5	exp. gew. Modell: β=6
CCC-neu	0,89**	exp. gew. Modell: β=0,1	6,5	exp. gew. Modell: β=0,8; 0,9 ;1,0
CHIC-alt	0,82*	exp. gew. Modell: β=0,3	4,3	exp. gew. Modell: β=3; 4
CHIC-neu	0,80	exp. gew. Modell: β=0,1	4,3	exp. gew. Modell: β=5
SE-ACA	0,48	exp. gew. Modell: β=0,1	7,2	exp. gew. Modell: β=2
SE-CCC	0,79	exp. gew. Modell: β=0,1	6,1	exp. gew. Modell: β=2
SE-CHIC	0,75	exp. gew. Modell: β=0,2	4,7	exp. gew. Modell: β=2

* $p<0,1$; ** $p<0,05$

Die Ergebnisse in Tabelle 5-11 bestätigen, daß die beiden Gütekriterien Pearson und MAA zu unterschiedlichen Befunden führen können. Zieht man zunächst den Korrelationskoeffizienten **Pearson** zur Beurteilung heran, ergibt sich eine Überlegenheit für die CCC-neu. Sie weist mit 0,89 eine signifikante Korrelation ($p<0,05$) zwischen den beobachteten und prognostizierten Marktanteilen der simulierten Kaufsituation aus. Noch schwach signifikant ($p<0,1$) ist die Korrelation bei der CHIC-alt mit einem Wert von 0,82. Damit kann die Überlegenheit dieser beiden Methoden gegenüber allen anderen Varianten indirekt als (schwach) signifikant interpretiert werden. Während die neue Berechnungsformel bei der CHIC von untergeordneter Rolle ist, führt sie bei der CCC zu einer Verbesserung der Prognosevalidität. Die ACA als drittes hybrides Verfahren schneidet bei diesem Gütekriterium deutlich schlechter ab. Die maximale Korrelation von 0,64 ist nicht nur niedriger als bei den beiden neuen hybriden Methoden, sondern zudem auch schlechter als bei den Self-Explicated-Varianten SE-CCC und SE-CHIC. Während die Unterschiede dieser beiden Self-Explicated-Methoden zur jeweils hybriden Vorgehensweise verhältnismäßig gering (bei der CCC-alt sogar gleich 0)

sind, fällt die Prognosevalidität der SE-ACA noch weiter ab. Sie weist mit 0,48 die insgesamt und mit deutlichem Abstand geringste Übereinstimmung zwischen beobachteten und prognostizierten Werten auf. Zwar führt die Hinzunahme der Paarvergleichsaufgabe zu einer Erhöhung von 0,48 auf 0,64, jedoch liegt auch dieser Wert noch weit unter allen anderen Verfahren.

Für ein Unternehmen ist die absolute Höhe des Marktanteils jedoch von größerer Bedeutung als der relative Vergleich zur den Wettbewerbern. Damit kommt der MAA als Maß für die Prognosevalidität eine größere Bedeutung zu. Die Ergebnisse diesbezüglich zeigen, daß sich die geringsten Abweichungen für die CHIC-alt und CHIC-neu ergeben. Überraschend gut ist auch die Prognose der Self-Explicated-Variante der CHIC. Mit Ausnahme der SE-ACA, die auch hier wiederum mit Abstand die geringste Prognosevalidität ausweist, ergeben sich für alle anderen Methoden vergleichbare Werte, aber durchweg höhere als bei der CHIC. Somit kann festgehalten werden, daß sich die relativen Ergebnisse der Methoden untereinander im Vergleich zu denen hinsichtlich des Korrelationskoeffizienten verändern.

Für die vorgeschlagene Modifikation der Berechnungsformel ergibt sich ebensowenig ein eindeutiges Bild. Während die neue Berechnungsformel bei der CHIC keinen bzw. einen sehr geringen Einfluß auf das Ergebnis nimmt, führt sie bei der CCC zu einer Erhöhung der MAA und dadurch zu einer Verschlechterung der Prognosevalidität. Wird die Prognosevalidität dagegen durch den Korrelationskoeffizienten nach Pearson quantifiziert, schneidet die CCC-neu besser ab als die CCC-alt. Somit sind die Befunde sowohl im Vergleich zwischen den beiden Verfahren als auch hinsichtlich alternativer Operationalisierungen der Gütekriterien widersprüchlich.

Vergleicht man diese Befunde der simulierten Marktanteile mit der individuellen Analyse auf Basis der individuellen Kendall t-Werte und der First-Hit-Rate fällt auf, daß sich auf aggregierter Ebene zumindest teilweise (schwach) signifikante Unterschiede ergeben. Demgegenüber treten bei beiden individuell gemessenen Gütekriterien keinerlei signifikante Unterschiede zwischen den Methoden auf.[339]

---

[339] Vgl. hierzu die Ergebnisse in Kapitel 5.3.1 und 5.3.2.

Ursache hierfür können die bereits diskutierten Reliabilitätsprobleme sein. Sofern Probanden im Rahmen der Holdout-Aufgabe anders geantwortet haben als in der eigentlichen Conjoint-Aufgabe, was die geringen Mittelwerte für die individuellen Kendalls τ-Werte vermuten lassen, wirken sich diese Verzerrungen der Validitätsergebnisse auf aggregierter Ebene geringer aus als auf individueller. Demzufolge können die Ergebnisse auch hinsichtlich der alternativen Gütekriterien voneinander abweichen, obwohl eigentlich das gleiche gemessen wird.[340]

In Tabelle 5-11 ist weiterhin aufgeführt, unter welcher Kaufverhaltensannahme die jeweils maximale Prognosevalidität ermittelt wurde. Für alle Methoden ist dies das exponentiell gewichtete Kaufverhaltensmodell. Unterschiede treten lediglich in der Ausprägung des Exponenten β auf. Während bei Anwendung des Korrelationskoeffizienten der Exponent bei allen Verfahren nur zwischen 0,1 und 0,3 schwankt, variiert dieser bei Bestimmung der MAA zwischen 0,8 (CCC-neu) und 6 (CCC-alt) bedeutend stärker. Die wesentlich höheren ß-Werte bei der MAA deuten darauf hin, daß eine zu starke Gleichverteilung der Kaufwahrscheinlichkeiten in ihrer absoluten Höhe, wie es insbesondere die BTL- und die Logit-Regel vornehmen, nicht den tatsächlich beobachteten Marktanteilen entspricht. Wie in Abschnitt 5.3.3.1 angeführt, ergeben sich sowohl für das BTL- als auch für das Logit-Modell tendenziell für alle betrachteten Alternativen sehr ähnliche Kaufwahrscheinlichkeiten.[341] Im vorliegenden Fall impliziert dies, daß für alle 5 betrachteten Angebote eine Kaufwahrscheinlichkeit und den daraus abgeleiteten Marktanteil von näherungsweise 20% ermittelt wird. Ein Vergleich der beobachteten Marktanteile in den drei Teilstichproben zeigt jedoch, daß die tatsächlichen Werte (vgl. Tabelle 5-12) deutlich von dieser Gleichverteilung zwischen den Angeboten abweichen. Da dieser Effekt mit steigendem Exponenten β beim exponentiell gewichteten Modell ausgeglichen werden kann, liefert diese Eigenschaft des Modells eine mögliche Erklärung für die deutlich höheren β-Werte bei Analyse der MAA gegenüber einer Operationalisierung durch den Korrelationskoeffizienten nach Pearson.

---

340 Vgl. Wittink/Bergestuen (1999), S. 11 f.

341 Es sei noch einmal darauf hingewiesen, daß das BTL-Model dem exponentiell gewichteten Modell mit einem Exponenten $\beta=1$ entspricht.

Tabelle 5-12: Beobachtete Marktanteile (in %) im Rahmen der simulierten Kaufentscheidung

Stichprobe	ACA	CCC	CHIC
Angebot 1	28,9	31,4	29,9
Angebot 2	13,7	21,9	19,7
Angebot 3	28,9	25,6	24,4
Angebot 4	7,2	7,4	9,4
Angebot 5	21,3	13,6	16,6

Für die „optimalen" β-Exponenten zeigt sich weiterhin, daß diese nicht nur über die verschiedenen Methoden schwanken, sondern auch bei jedem Verfahren selbst zwischen den beiden Gütekriterien Pearson und MAA variieren. In einem weiteren Schritt der Auswertung wird deshalb untersucht, inwieweit die Ergebnisse hinsichtlich der Prognosegüte von den unterstellten Kaufverhaltensannahmen abhängig sind.

5.3.3.3 Einfluß der Kaufverhaltensannahme auf die Prognosevalidität

Für den Vergleich der jeweils maximalen Prognosegüte für die alternativen Methoden wurde jeweils die Kaufverhaltensannahme herangezogen, bei der sich die höchste Übereinstimmung zwischen prognostizierten und empirisch beobachteten Werten der simulierten Marktanteile ergab. Damit ist jedoch zunächst nur ein Teil der Ergebnisse berichtet worden. Ebenso von zentraler Bedeutung ist, inwieweit die Prognoseergebnisse von der Wahl der Kaufverhaltensannahme beeinflußt werden. Während kein Zweifel darüber besteht, *das* die Kaufverhaltensannahme einen Einfluß hat,[342] ist noch unklar *wie stark* die Prognoseergebnisse davon abhängen. Insbesondere ist für den Validitätsvergleich der Methoden entscheidend, ob das *relative Abschneiden* der Verfahren durch die Wahl der Kaufverhaltensannahme verändert wird. Diese Fragestellung ist bisher in der Literatur vernachlässigt worden, denn häufig werden nur die Ergebnisse unter Annahme einer Kaufverhaltensannahme berichtet.[343] In der vorliegenden Untersuchung wird daher die Sensibilität der Prognoseergebnisse hinsichtlich der Kaufverhaltensannahme explizit analysiert.

---

342 Vgl. Green/Krieger (1988), S. 115.

343 Vgl. Agarwal/Green (1991), S. 144; Huber et al. (1993), S. 109; Pullman/Dodson/Moore (1999), S. 131 f.

In den beiden folgenden Tabellen sind die ermittelten Werte für den Korrelationskoeffizienten Pearson und die MAA bei allen 30 verschiedenen Modellierungen der Kaufverhaltensannahme aufgeführt. Die jeweils optimalen Werte aus Tabelle 5-11 im vorangegangenen Kapitel sind hervorgehoben. Weiterhin sind am Ende der Tabelle der Mittelwert, sowie Maximum und Minimum des jeweiligen Gütekriteriums angegeben. Die Analyse dieser Ergebnisse kann Aufschluß darüber geben, wie stark die erzielte Prognosevalidität auf aggregierter Ebene von der Wahl der Kaufverhaltensannahme beeinflußt wird.

Tabelle 5-13: Korrelationskoeffizienten (Pearson) zwischen beobachteten und prognostizierten Werten der simulierten Marktanteile bei unterschiedlichen Kaufverhaltensmodellen

Pearson		ACA	CCC-alt	CCC-neu	CHIC-alt	CHIC-neu	SE-ACA	SE-CCC	SE-CHIC
Kaufverhaltensmodelle	FC a)	0,43	0,67	0,63	0,65	0,53	0,43	0,67	0,67
	Logit	0,62	0,74	0,69	0,75	0,74	0,43	0,70	0,73
	β=0,1	0,61	**0,79**	**0,89****	0,79	**0,80**	**0,48**	**0,79**	0,72
	β=0,2	**0,64**	0,76	0,81*	0,81*	0,78	0,42	0,70	**0,75**
	β=0,3	0,63	0,78	0,79	**0,82***	0,77	0,43	0,74	0,75
	β=0,4	0,63	0,76	0,77	0,78	0,77	0,39	0,72	0,74
	β=0,5	0,62	0,75	0,74	0,78	0,76	0,41	0,72	0,75
	β=0,6	0,59	0,76	0,72	0,77	0,75	0,40	0,70	0,74
	β=0,7	0,58	0,75	0,69	0,76	0,75	0,39	0,69	0,71
	β=0,8	0,58	0,75	0,68	0,77	0,74	0,39	0,70	0,71
	β=0,9	0,58	0,73	0,67	0,75	0,74	0,39	0,69	0,71
	β=1 b)	0,58	0,73	0,66	0,75	0,72	0,39	0,68	0,70
	β=2	0,55	0,68	0,59	0,70	0,67	0,36	0,64	0,66
	β=3	0,53	0,65	0,55	0,68	0,65	0,36	0,62	0,64
	β=4	0,51	0,64	0,55	0,68	0,65	0,37	0,61	0,64
	β=5	0,51	0,64	0,55	0,68	0,64	0,38	0,61	0,64
	β=6	0,50	0,65	0,55	0,68	0,64	0,39	0,61	0,65
	β=7	0,50	0,65	0,55	0,68	0,65	0,40	0,62	0,65
	β=8	0,49	0,65	0,55	0,68	0,65	0,40	0,62	0,66
	β=9	0,49	0,66	0,55	0,68	0,65	0,41	0,62	0,66
	β=10	0,49	0,66	0,56	0,69	0,65	0,41	0,63	0,66
	β=20	0,47	0,68	0,58	0,68	0,64	0,42	0,65	0,67
	β=30	0,46	0,68	0,59	0,68	0,64	0,43	0,66	0,67
	β=40	0,46	0,68	0,60	0,68	0,64	0,44	0,67	0,67
	β=50	0,45	0,69	0,60	0,67	0,63	0,43	0,67	0,67
	β=60	0,45	0,69	0,61	0,67	0,63	0,43	0,67	0,67
	β=70	0,45	0,69	0,61	0,67	0,63	0,44	0,67	0,67
	β=80	0,45	0,68	0,61	0,67	0,63	0,43	0,67	0,67
	β=90	0,44	0,68	0,61	0,67	0,63	0,43	0,67	0,67
	β=100	0,44	0,68	0,61	0,67	0,63	0,44	0,67	0,67
MW		0,52	0,70	0,64	0,71	0,68	0,41	0,67	0,69
MAX		0,64	0,79	0,89	0,82	0,80	0,48	0,79	0,75
MIN		0,43	0,64	0,55	0,65	0,53	0,36	0,61	0,64

* $p<0,1$;** $p<0,05$

a) First-Choice-Model

b) entspricht dem BTL-Model

Tabelle 5-14: Mittlere absolute Abweichungen (MAA in %) zwischen den beobachteten und prognostizierten Werten der simulierten Marktanteile bei unterschiedlichen Kaufverhaltensmodellen

MAA		ACA	CCC-alt	CCC-neu	CHIC-alt	CHIC-neu	SE-ACA	SE-CCC	SE-CHIC
Kaufverhaltensmodelle	**FC** a)	12,0	10,7	7,4	12,5	9,2	18,3	15,3	15,6
	Logit	7,4	7,0	7,1	5,5	5,5	7,5	6,9	5,4
	β=0,1	7,6	7,4	7,4	5,6	5,6	7,6	7,4	5,6
	β=0,2	7,4	7,2	7,3	5,5	5,6	7,6	7,2	5,5
	β=0,3	7,4	7,0	7,1	5,4	5,5	7,5	7,0	5,4
	β=0,4	7,3	6,9	7,0	5,4	5,4	7,5	6,7	5,4
	β=0,5	7,2	6,7	6,8	5,3	5,4	7,5	6,5	5,3
	β=0,6	7,1	6,5	6,7	5,3	5,3	7,4	6,5	5,2
	β=0,7	7,1	6,4	6,6	5,2	5,3	7,4	6,4	5,2
	β=0,8	7,0	6,3	**6,5**	5,1	5,2	7,4	6,3	5,1
	β=0,9	6,9	6,3	**6,5**	5,1	5,2	7,4	6,3	5,0
	β=1 b)	6,9	6,2	**6,5**	5,0	5,1	7,3	6,3	5,0
	β=2	6,3	6,0	6,7	4,6	4,9	**7,2**	**6,1**	**4,7**
	β=3	**5,9**	5,9	6,9	4,3	4,7	7,6	6,1	4,9
	β=4	6,5	5,8	7,0	**4,3**	4,5	8,6	6,6	5,1
	β=5	7,0	5,6	7,0	4,7	**4,3**	9,4	6,9	6,0
	β=6	7,5	**5,5**	7,0	5,3	4,7	10,2	7,4	7,1
	β=7	7,9	6,0	6,9	6,0	5,1	11,1	8,1	8,1
	β=8	8,3	6,4	6,9	6,6	5,6	12,0	8,7	8,9
	β=9	8,7	6,8	6,9	7,2	6,0	12,8	9,2	9,7
	β=10	9,0	7,1	6,8	7,7	6,4	13,5	9,7	10,3
	β=20	10,5	9,4	7,2	10,4	8,5	16,6	12,8	13,5
	β=30	11,1	10,2	7,6	11,4	9,2	17,4	13,9	14,6
	β=40	11,4	10,5	7,8	11,8	9,5	17,7	14,4	15,0
	β=50	11,5	10,7	7,9	12,0	9,7	17,9	14,6	15,2
	β=60	11,7	10,8	7,9	12,1	9,9	18,0	14,8	15,3
	β=70	11,7	10,8	7,9	12,2	9,9	18,0	14,9	15,4
	β=80	11,8	10,9	7,9	12,3	9,9	18,1	15,0	15,4
	β=90	11,8	10,9	7,9	12,3	10,0	18,2	15,0	15,5
	β=100	11,8	10,9	7,9	12,3	10,0	18,2	15,1	15,5
MW		8,7	7,8	7,2	7,6	6,7	11,8	9,6	9,1
MAX		12,0	10,9	7,9	12,5	10,0	18,3	15,3	15,6
MIN		5,9	5,5	6,5	4,3	4,3	7,2	6,1	4,7

a) First-Choice-Model

b) entspricht dem BTL-Model

Die Verläufe beider Gütekriterien (vgl. Tabelle 5-13 und 5-14) zeigen bei allen acht Methoden, daß die Höhe der Übereinstimmung zwischen prognostizierten und beobachteten Werten je nach angenommenem Kaufverhaltensmodell variiert. Für den Korrelationskoeffizienten weist die CCC-neu mit einer Schwankungsbreite von 0,55 bis 0,89 die größten Unterschiede zwischen den Ergebnissen alternativer Kaufverhaltensregeln auf. Im Gegensatz dazu variiert die MAA bei diesem Verfahren nur zwischen 6,5 und 7,2 und weist die insgesamt geringsten Schwankungen bei diesem Gütekriterium auf. Dies bestätigt erneut, daß die beiden Gütekriterien zu unterschiedlichen Ergebnisinterpretationen führen können. Ein möglicher Grund hierfür kann darin liegen, daß sich Änderungen der Kaufverhaltensannahme stärker auf die relativen Unterschiede zwischen den ermittelten Marktanteilen (gemessen durch Pearson) als auf deren absolute Höhe (gemessen durch MAA) auswirken. Bei allen anderen Methoden schwankt die MAA deutlich stärker, wobei die Abweichungen bei den Self-Explicated-Methoden mit 11,1% (SE-ACA) bis auf das achtfache (gegenüber der CCC-neu mit 1,4%) ansteigen. Darin wird deutlich, daß die Wahl der Kaufverhaltensannahme die Ausprägung der Prognosevalidität maßgeblich determiniert.

Vergleicht man die Ergebnisse über die verschiedenen Kaufverhaltensmodelle hinweg, ist außerdem festzustellen, daß auch das relative Abschneiden der Methoden unterschiedlich ausfällt. Um diese Divergenzen zu quantifizieren, werden die 8 Methoden zunächst entsprechend der Höhe jeweils bei beiden Gütemaßen in eine Rangfolge gebracht. So wird z.B. der CCC-neu (SE-ACA) unter Annahme des First-Choice-Model der Rangplatz 1 (8) zugeordnet, da sie im Vergleich zu allen anderen Methoden mit 7,4% (18,3%) die geringste (höchste) MAA aufweist (vgl. Tabelle 5-14). Über den Rangkorrelationskoeffizienten nach Kendall kann in einem nächsten Schritt überprüft werden, inwieweit diese Rangfolgen der Methoden bei den unterschiedlichen Kaufverhaltensmodellen übereinstimmen. Für die beiden Gütekriterien wurden entsprechend der 30 angewendeten Modelle insgesamt 435 Korrelationskoeffizienten (nach Kendall) berechnet.[344] Nimmt dieser Korrelationskoeffizient den Wert 1 an, so führen zwei unterschiedliche Kaufverhaltensannahmen zum identischen relativen Abschneiden der 8 Methoden. Die Vergleiche zwischen allen Kaufverhaltensannahmen zeigen, daß dies nur in 6%

[344] Die Anzahl dieser paarweisen Vergleiche ergibt sich aus n(n-1)/2, wobei n für die Anzahl der angewendeten Kaufverhaltensmodelle (=30) steht.

aller Fälle zutrifft, sofern die Prognosegüte auf Basis der Korrelation nach Pearson quantifiziert wird. Dient die MAA als Gütekriterium, liegt der Prozentsatz mit 11% übereinstimmender Befunde nicht wesentlich höher.

Wie stark die Ergebnisse und die Unterschiede zwischen den Verfahren in ihrer absoluten Höhe differieren, bleibt bei der Bestimmung der Rangplätze allerdings unberücksichtigt. Diese Überlegung soll an einem vereinfachten Beispiel verdeutlicht werden (vgl. hierzu Tabelle 5-13). Für β=0,1 ergibt sich für die CHIC-neu (Pearson = 0,80) ein niedrigerer Rangplatz als für die CHIC-alt (Pearson = 0,79). Das relative Abschneiden (gemessen an den Rangplätzen) kehrt sich in der nächsten Zeile (β=0,2) genau um. Wird die Übereinstimmung nun anhand des Rangkorrelationskoeffizienten quantifiziert, bleibt unberücksichtigt, daß sich die absolute Höhe der Gütemaße bei beiden Verfahren lediglich um 0,02 verändert hat. Die Integration dieser Unterschiede in die Betrachtung erfolgte in einem zweiten Schritt über die Berechnung der Korrelationen zwischen dem relativen Abschneiden der Verfahren bei unterschiedlichen Kaufverhaltensannahmen mit dem Korrelationskoeffizient nach Pearson. Hierbei ergeben sich in 63% (Pearson) bzw. 36% (MAA) der jeweils 435 Vergleiche Korrelationen über 0,9. Der Grad der Übereinstimmung zwischen den alternativen Kaufverhaltensmodellen fällt bei dieser Vorgehensweise z.T. wesentlich höher aus. Dennoch kann für die vorliegende Untersuchung wie erwartet festgehalten werden, daß alternative Kaufverhaltensannahmen zu unterschiedlichen Prognoseergebnissen führen können. Das in diesem Zusammenhang neue und wichtigere Ergebnis dieser Analyse ist, daß die Wahl der Kaufverhaltensannahme auch den relativen Verfahrensvergleich beeinflußt.

Welche Kaufverhaltensannahme zur höchsten Prognosevalidität führt und welche weniger geeignet ist, zeigt sich auf Basis der bisherigen Ergebnisse in der Tendenz über alle Methoden einheitlich. Bereits in Kapitel 5.3.3.2 bei der Gegenüberstellung der maximalen Werte für die Prognosevalidität wurde diskutiert, daß bei allen Methoden mit dem exponentiell gewichteten Modell die höchste Übereinstimmung erzielt wird. Auch die Wertebereiche für den gewählten Exponenten β sind z.T. über die verschiedenen Methoden vergleichbar. Einheitlich über alle Methoden ist ebenfalls das schlechte Abschneiden des First-Choice-Model. Dieser Befund entspricht den Ergebnissen von Huber et al. (1993). Die Autoren be-

richten, daß sich die First-Choice-Regel auch bei ihrer Studie mit Kühlschränken als Untersuchungsgegenstand als nicht geeignet zur Ableitung der Kaufwahrscheinlichkeiten erwies.[345] Dennoch widersprechen diese Befunde den erwarteten Vorteilen des Modells. Gerade für langlebige Gebrauchsgüter, erscheint das First-Choice-Model zumindest aus theoretischer Sicht als sehr geeignet.[346] Eine Erklärung für die vorliegenden Befunde kann darin liegen, daß für einen Großteil der Probanden die Produktalternativen des Holdout-Sets in ihrer Attraktivität, gemessen durch die geschätzten Gesamtnutzen, sehr ähnlich sind. Wie in Kapitel 5.3.3.1 erläutert, kann das First-Choice-Model in diesen Fällen zu verzerrten Prognoseergebnissen führen.

5.3.4 Zusammenfassung der Ergebnisse hinsichtlich der Prognosevalidität

Die Prüfung der Prognosevalidität fußt auf Präferenzdaten, die über eine zusätzliche Holdout-Aufgabe erhoben wurden. Diese sind auf der Grundlage der geschätzten Teilnutzenwerte prognostiziert und mit den empirisch beobachteten Werten verglichen worden. Die Übereinstimmung wurde dabei anhand unterschiedlicher Gütekriterien gemessen. Als zentrale Ergebnisse können festgehalten werden:

1. Hinsichtlich der First-Hit-Rate weist die ACA, CHIC-neu und die CCC-alt die beste Prognosevalidität auf. Sie führen alle drei zu sehr ähnlichen Werten für dieses Gütekriterium. Alle anderen Methoden schneiden etwas schlechter ab und variieren nur geringfügig untereinander. Insgesamt erweisen sich aber alle Unterschiede zwischen den Methoden als nicht signifikant.
2. Analog zur First-Hit-Rate ergeben sich auch bei einem Vergleich der individuellen Rangkorrelationen (Kendall) keine signifikanten Unterschiede zwischen den 8 untersuchten Methoden. Die Mittelwerte über alle Probanden variieren mit Werten von 0,26 (CHIC-alt) und 0,33 (CCC-alt bzw. CCC-neu) nur unbedeutend. Auffallend ist allerdings, daß die Rangkorrelationen in ihrer absoluten Höhe unerwartet niedrig ausfallen.
3. Die Übereinstimmung zwischen beobachteten und prognostizierten Werten für die simulierten Marktanteile wurde sowohl durch die Korrelation (Pearson)

---

[345] Vgl. Huber et al. (1993), S. 109.
[346] Vgl. Gutsche (1995), S. 142 f. und Green/Krieger (1988), S. 118.

als auch durch die mittlere absolute Abweichung (MAA) gemessen. Diese beiden Gütekriterien führen zu unterschiedlichen Befunden. Hinsichtlich des Korrelationskoeffizienten nach Pearson zeigt sich eine Überlegenheit für die CCC-neu. Während alle anderen Varianten der beiden neuen Verfahren CCC und CHIC, insbesondere auch deren Self-Explicated-Phasen niedrigere aber untereinander vergleichbare Ergebnisse aufweisen, schneiden die ACA und die SE-ACA schlechter ab. Eine höhere Aussagekraft liefert allerdings die MAA. Bei beiden Varianten der CHIC tritt mit 4,3% die geringste mittlere Abweichung auf, gefolgt von der SE-CHIC mit einer MAA von 4,7%. Alle anderen Verfahren schneiden schlechter ab, wobei die Unterschiede (mit Ausnahme der SE-ACA) auch hier vergleichsweise gering sind.

4. Die Prognoseergebnisse auf aggregierter Ebene, gemessen durch simulierte Marktanteile, werden maßgeblich davon beeinflußt, welches Kaufverhaltensmodell zur Transformation der geschätzten Gesamtnutzenwerte in Kaufwahrscheinlichkeiten bzw. Marktanteile angenommen wird. Sowohl die Prognosevalidität der einzelnen Methoden in ihrer absoluten Höhe als auch das relative Abschneiden der Verfahren variiert in Abhängigkeit der Kaufverhaltensannahme.
5. Auf individueller Ebene (First-Hit-Rate und individuelle Kendalls τ-Werte) treten zwischen den untersuchten Methoden nur geringe, überwiegend nicht signifikante Unterschiede auf. Somit wirkt sich die auf individueller Ebene nachgewiesene diskriminierende Validität nicht auf die Prognosegüte aus. Wird dagegen die Prognosevalidität auf aggregierter Ebene (simulierte Marktanteile) überprüft, treten Unterschiede zwischen den Methoden auf, die sich jedoch für die CCC-neu und CHIC-alt als (schwach) signifikant erweisen.
6. Über alle Operationalisierungen hinweg variieren die Befunde hinsichtlich der alternativen Schätzverfahren für die CCC und die CHIC. So ist weder zwischen den Verfahren eine eindeutige Tendenz erkennbar, noch innerhalb einer Variante bei einem Vergleich der unterschiedlichen Gütemaße. Die Modifikation der Berechnungsformel führt teilweise zu einer Erhöhung oder zu einer Reduzierung der Prognosevalidität sowohl bei der CCC als auch bei der CHIC. Die Ergebnisse der Prognosevalidität lassen damit keine eindeutige Aussage über den Effekt der neuen Berechnungsformel zu.

5.4 Externe Validität

Alle bisherigen Validitätsprüfungen basieren auf Informationen, die im Rahmen der Interviews gewonnen wurden. Mit Blick auf die externe Validität der verschiedenen Methoden gilt es nun, in einem letzten Schritt zu testen, ob die bisherigen Befunde Bestand haben, wenn die auf Grundlage der Befragung prognostizierten Marktanteile mit realen Marktanteilen und damit mit von den befragten Stichproben unabhängigen Daten verglichen werden. Als Vergleichsgröße standen die realen Marktanteilsdaten des GfK-Handelspanel für das Jahr 1998 zur Verfügung.[347]

Obwohl die Prüfung der externen Validität durch einen Vergleich der Prognoseergebnisse mit realen Marktdaten das für die Praxis relevanteste Prüfkriterium für die Präferenzmeßmethoden darstellt, liegen bisher relativ wenige solcher Untersuchungen vor.[348] Wenn auch kein Zweifel über die Notwendigkeit von Validitätsvergleichen in Zusammenhang mit realen Marktanteilen besteht,[349] überrascht dieses bisherige Defizit solcher Studien in der Literatur kaum. Die Schwierigkeit liegt maßgeblich in der Übertragung der Marktanteilsprognosen auf Basis eines Conjoint-Interviews auf die reale Kauf- bzw. Marktsituation. Obgleich Entscheidungen im Rahmen eines Conjoint-Interviews immer unter Annahme einer tatsächlichen Kaufentscheidung getroffen werden, ist nicht sichergestellt, daß diese dem realen Kaufverhalten entsprechen. Erschwert wird die Übertragbarkeit durch weitere Einflüsse auf das Verhalten in der realen Kaufsituation. So hängt die tatsächliche Kaufentscheidung unter anderem von der Wirkung verschiedener Marketing-Mix-Instrumente wie Werbung und Distribution ab.[350] Als weitere Einflußgrößen führen Wittink/Bergestuen (1999) die Dynamik des Marktgeschehens oder das Auftreten zusätzlicher Marktteilnehmer an.[351] Während die Markt-

---

347 Für die Bereitstellung der Daten sei dem GfK-Panelservice, namentlich den Herren H.-J. Finck und Friedemann Stöckle an dieser Stelle Dank ausgesprochen.

348 Einen Überblick über Studien, bei denen die externe Validität überprüft wurde, geben Levin et al. (1983), Louviere (1988) und Green/Srinivasan (1990), wobei die Autoren die externe Validität hierbei nicht ausschließlich auf die Prognose realer Marktanteile beziehen, sondern beispielsweise auch um die Prognose tatsächlichen Entscheidungsverhaltens erweitern. Diese erweiterte Definition der externen Validität entspricht der im Rahmen dieser Arbeit vorgenommenen Abgrenzung der Gütekriterien (vgl. hierzu Kapitel 2.2.2.1).

349 Vgl. Green/Srinivasan (1990), S. 13.

350 Vgl. Srinivasan/deMaCarty (1998), S. 1.

351 Vgl. Wittink/Bergestuen (1999), S. 8.

anteilsprognosen im Rahmen eines Conjoint-Interviews auf der Annahme basieren, daß die Probanden alle Produkte kennen und zudem alle in gleicher Weise verfügbar sind, kann beides in der realen Marktsituation über die Produkte variieren.[352]

Um den Einfluß dieser zusätzlichen Variablen ausklammern zu können, schlagen Srinivasan/deMaCarty (1998) vor, die Validierung der Prognoseergebnisse nicht auf Basis absoluter Verkaufszahlen vorzunehmen, sondern „nur" deren Verhältnisse hinsichtlich verschiedener Produkte zu prognostizieren. Die Autoren betrachten hierfür die Situation einer Unternehmung, die mehrere Produkte in einer Produktkategorie anbietet. Unter Annahme eines vergleichbaren Zeithorizonts und gleich hohen Marketingaufwendungen für die verschiedenen Produkte, sind Daten wie z.B. Werbeaufwendungen, Distributionsquoten etc. für die Validierung nicht mehr erforderlich.[353] Übertragen auf die vorliegende Studie impliziert dieses Vorgehen, daß relative Verkaufszahlen verschiedener Modelle von Tischkühlschränken *eines* Herstellers zu prognostizieren sind. Da jedoch die Absatzzahlen der verschiedenen Modelle in der Produktkategorie „Tischkühlschrank" für die einzelnen Anbieter nicht vorlagen, konnte das vorgeschlagene Verfahren zur Validierung nicht angewendet werden. Dennoch wird die Argumentation von Srinivasan/deMaCarty (1998) in der vorliegenden Validitätsprüfung aufgegriffen. Die Operationalisierung der Prognosevalidität als auch der externen Validität durch den Korrelationskoeffizient nach Pearson zwischen geschätzten und beobachteten Marktanteilen mißt ebenfalls die relative Marktposition der verschiedenen Anbieter.

Um dennoch die externe Validität der untersuchten Verfahren überprüfen zu können, werden bei der Prognose realer Marktanteile entsprechende Kontextvariablen wie Marketing-Mix-Instrumente integriert. Das hierfür angewendete Marktanteilsmodell wird im folgenden Abschnitt erläutert.

---

352 Vgl. Srinivasan/deMaCarty (1998), S. 3.
353 Vgl. Srinivasan/deMaCarty (1998), S. 4 ff.

5.4.1 Marktanteilsmodell

Grundlage für die Marktanteilsprognose war folgendes in der Literatur und Praxis vielfach (in dieser oder leicht modifizierter Form) angewandte Modell[354]:

$$(5\text{-}5) \quad MA_m = DG_m * \sum_{i \in I} \frac{ER_{im} * BG_{im}}{I} \qquad (i \in I;\ m \in M)$$

wobei:

MA_m: Prognostizierter Marktanteil für das m-te Produkt,

DG_m: Distributionsgrad des m-ten Produkts,

ER_{im}: Erstkaufrate für das m-te Produkt des i-ten Probanden,

BG_{im}: Bekanntheitsgrad des m-ten Produkts beim i-ten Probanden,

I: Indexmenge der Probanden,

M: Indexmenge der Alternativen.

In seiner Grundform berücksichtigt dieses Marktanteilsmodell zusätzlich die Wiederkaufrate und die Kaufintensität für das jeweilige Produkt. Da in der vorliegenden Studie mit „Kühlschränken" als Untersuchungsgegenstand ein langlebiges Gebrauchsgut betrachtet wurde, kommt diesen beiden Größen eine untergeordnete Rolle zu und wurden daher nicht in das Modell integriert. Eine Auswertung der im Rahmen der Befragung erhobenen Informationen zur Wiederkaufrate bzw. Kaufintensität bestätigt diese Vorgehensweise. In der vorliegenden Studie gaben lediglich 10,4% der 725 Probanden an, schon ein zweites Mal einen Kühlschrank der gleichen Marke gekauft zu haben. Häufiger als zweimal wurde eine Marke von keinem der Probanden gekauft.

Zur Umsetzung des Modells müssen die auf der rechten Seite der Gleichung (5-5) angegebenen Größen empirisch bestimmt werden. Für die **Erstkaufrate** wurde analog zur Vorgehensweise in Kapitel 5.3 die Kaufwahrscheinlichkeit für jede

---

[354] Vgl. Silk/Urban (1978), S. 177; Erichson (1979); S. 261 f.; Hammann/Erichson (1994), S. 378 ff.; Gaul/Baier/Aperis (1996), S. 210.

der 4 betrachteten Marken abgeleitet.[355] Hierfür wurden zunächst für alle verschiedenen Modelle von Tischkühlschränken der jeweiligen Hersteller die Gesamtnutzenwerte berechnet. Der jeweils individuell höchste Gesamtnutzen einer Marke diente zur Bestimmung der Erstkaufrate.[356] Die Transformation der Gesamtnutzenwerte in Kaufwahrscheinlichkeiten wurde ebenfalls unter Annahme der verschiedenen Kaufverhaltensmodelle vorgenommenen. Die **Bekanntheitsgrade** für die verschiedenen Marken wurden im Rahmen des Interviews individuell durch direkte Befragung erhoben.[357] Für die **Distributionsgrade** standen die aggregierten Daten des GfK-Handelspanel aller 4 untersuchten Marken für das Jahr 1998 zur Verfügung.[358]

Die Messung der externen Validität wird analog zur Prognosevalidität anhand des Korrelationskoeffizienten nach Pearson und der MAA vorgenommen. Dabei werden zunächst die Ergebnisse der maximalen externen Validität dargestellt. In einem zweiten Schritt wird geprüft, wie die Annahmen über das Kaufverhalten die Wertebereiche der beiden Gütemaße beeinflussen.

5.4.2 Maximale externe Validität der verschiedenen Methoden

Die maximalen Übereinstimmungen der verschiedenen Methoden zwischen prognostizierten und beobachteten Werten für die realen Marktanteile sind in Tabelle 5-15 wiedergegeben.

---

355 Aus Gründen der Vergleichbarkeit zur Prognosevalidität sollten jeweils nur die Marken betrachtet werden, die auch in der simulierten Kaufentscheidung enthalten waren. Da eine dieser Marken jedoch nicht im GfK-Handelspanel berücksichtigt ist, wird die Marktanteilsprognose an dieser Stelle nur für die verbleibenden 4 Marken des Holdout-Sets vorgenommen.

356 Diese Vorgehensweise war notwendig, da die Marktanteilsdaten nur für Tischkühlschränke insgesamt und nicht für die verschiedenen Modelle differenziert vorlagen.

357 Um Verzerrungen des Verfahrensvergleichs aufgrund unterschiedlicher Bekanntheitsgrade zwischen den befragten Teilstichproben zu vermeiden, werden die Bekanntheitsgrade jeweils über die drei Teilstichproben gemittelt.

358 Dabei muß berücksichtigt werden, daß es sich um die Distributionsgrade für die gesamte Bundesrepublik handelte, wohingegen die Daten der vorliegenden Befragung mit der regionalen Begrenzung auf die drei Städte Hamburg, Frankfurt/Main und München erhoben wurden. Die direkte Übertragbarkeit der GfK-Daten auf die vorliegende Untersuchung ist insofern mit Einschränkungen in der Ergebnisinterpretation verbunden.

Tabelle 5-15: Maximale Übereinstimmungen zwischen beobachteten und prognostizierten Werten der realen Marktanteile gemessen durch Pearson und mittlere absolute Abweichung (MAA)

Methode	Pearson	Kaufverhaltensmodell	MAA in %	Kaufverhaltensmodell
ACA	0,87	exp. gew. Modell: β=4;5;6;7	4,8	exp. gew. Modell: β=5
CCC-alt	0,88	exp. gew. Modell: β=5;6;7;8;9;10	4,4	exp. gew. Modell: β=5
CCC-neu	0,94*	exp. gew. Modell: β=10	3,8	exp. gew. Modell: β=6
CHIC-alt	0,88	exp. gew. Modell: β=6;7;8;9;10	4,5	exp. gew. Modell: β=5
CHIC-neu	0,93*	exp. gew. Modell: β=8;9;10	3,9	exp. gew. Modell: β=6
SE-ACA	0,82	exp. gew. Modell: β=0,8;0,9;1;2;3	5,4	exp. gew. Modell: β=3
SE-CCC	0,83	exp. gew. Modell: β=3;4;5;6;7;8	5,0	exp. gew. Modell: β=5
SE-CHIC	0,84	exp. gew. Modell: β=3;4;5;6;7	5,0	exp. gew. Modell: β=4

* $p<0,1$

Im Vergleich zu den Ergebnissen hinsichtlich der simulierten Marktanteile in Kapitel 5.3 fällt zunächst auf, daß insbesondere die Korrelationskoeffizienten höher ausfallen. Dies impliziert indes nicht ausschließlich, daß hier eine höhere Prognosegüte erzielt wird, sondern wird zumindest teilweise durch die geringere Anzahl an Beobachtungen (4 statt vorher 5 Marken) bedingt. Die Werte der MAA deuten darauf hin, daß die Prognose der realen Marktanteile verhältnismäßig gut gelingt. Bezüglich der beiden Gütekriterien (Korrelation nach Pearson und MAA) ergibt sich insgesamt eine Überlegenheit der beiden neuen Verfahren, wenn jeweils die neue Berechnungsformel angewendet wird. Lediglich bei diesen beiden Varianten ist eine (schwach) signifikante Korrelation zwischen den prognostizierten und beobachteten Marktanteilen festzustellen. Hinsichtlich der MAA weisen sie ebenfalls die höchste Übereinstimmung auf. Das Abschneiden der anderen Verfahren ist hinsichtlich beider Kriterien vergleichbar. Die ACA und die beiden neuen Methoden mit der „alten" Berechnungsformel führen untereinander zu ähnlichen Ergebnissen. Alle Self-Explicated-Methoden weisen eine geringere externe Vali-

dität auf als alle hybriden Verfahren. Analog zu allen bisherigen Befunden fällt diese Unterlegenheit überraschend gering aus.

Tabelle 5-15 belegt weiter, daß wiederum bei allen Methoden das exponentiell gewichtete Kaufverhaltensmodell am besten geeignet ist, die Gesamtnutzenwerte in Kaufwahrscheinlichkeiten zu transformieren. Im Vergleich zu den Befunden der Prognosevalidität sind hier jedoch wesentlich höhere Werte für den Exponenten β anzunehmen. Eine Analyse der realen Marktanteile gibt eine Erklärung für diesen Unterschied. Die realen Marktanteile der untersuchten Marken differieren bei den realen Marktanteilsdaten wesentlich stärker als die im Rahmen der simulierten Kaufsituation beobachteten (vgl. hierzu Tabelle 5-12).[359] Wie in Kapitel 5.3.3.1 diskutiert, fallen mit steigendem Exponenten β die Unterschiede zwischen den prognostizierten Marktanteilen bei der exponentiell gewichteten Kaufwahrscheinlichkeit größer aus. Den größeren Unterschieden in den realen Marktanteilen wird somit durch höhere Werte für β Rechnung getragen. Inwieweit die Ergebnisse der externen Validität von der jeweiligen Kaufverhaltensannahme abhängen, wird im zweiten Schritt durch Auswertung aller angewendeten Modelle getestet.

5.4.3 Einfluß der Kaufverhaltensannahme auf die externe Validität

Entsprechend der Vorgehensweise bei der Prognosevalidität in Abschnitt 5.3.3.3 wird auch an dieser Stelle überprüft, inwieweit die Ergebnisse hinsichtlich der externen Validität, insbesondere das relative Abschneiden der verschiedenen Methoden, davon abhängen, welche Kaufverhaltensannahme zur Transformation der Gesamtnutzenwerte herangezogen wird. Der Kaufverhaltensannahme spielt bei der Prognose realer Marktanteile eine zentrale Rolle. Vor diesem Hintergrund soll nun überprüft werden, inwieweit die Ergebnisse des Methodenvergleichs davon beeinflußt werden. In den beiden folgenden Tabellen (vgl. Tabelle 5-16 und Tabelle 5-17) sind die Ergebnisse der beiden Gütekriterien Pearson und MAA dargestellt.

---

[359] Aus Gründen der Geheimhaltung können die realen Marktanteile der untersuchten Marken nicht berichtet werden.

Tabelle 5-16: Korrelationskoeffizienten (Pearson) zwischen beobachteten und prognostizierten Werten der realen Marktanteile bei unterschiedlichen Kaufverhaltensmodellen

	Methode	**ACA**	**CCC-alt**	**CCC-neu**	**CHIC-alt**	**CHIC-neu**	**SE-ACA**	**SE-CCC**	**SE-CHIC**
Kaufverhaltensmodelle	**FC** a)	0,77	0,77	0,84	0,81	0,85	0,61	0,66	0,68
	Logit	0,82	0,82	0,82	0,82	0,82	0,81	0,81	0,81
	β=0,1	0,81	0,81	0,81	0,80	0,81	0,81	0,80	0,80
	β=0,2	0,81	0,81	0,81	0,81	0,81	0,81	0,80	0,81
	β=0,3	0,81	0,81	0,81	0,81	0,81	0,81	0,80	0,81
	β=0,4	0,81	0,81	0,82	0,81	0,81	0,81	0,81	0,81
	β=0,5	0,82	0,81	0,82	0,81	0,82	0,81	0,81	0,81
	β=0,6	0,82	0,82	0,82	0,82	0,82	0,81	0,81	0,81
	β=0,7	0,82	0,82	0,82	0,82	0,82	0,81	0,81	0,81
	β=0,8	0,82	0,82	0,83	0,82	0,82	**0,82**	0,81	0,82
	β=0,9	0,82	0,82	0,83	0,82	0,83	**0,82**	0,81	0,82
	β=1 b)	0,83	0,82	0,83	0,82	0,83	**0,82**	0,81	0,82
	β=2	0,84	0,84	0,86	0,84	0,85	**0,82**	0,82	0,83
	β=3	0,86	0,86	0,88	0,86	0,87	**0,82**	**0,83**	**0,84**
	β=4	**0,87**	0,87	0,89	0,87	0,89	0,81	**0,83**	**0,84**
	β=5	**0,87**	**0,88**	0,91*	0,87	0,90*	0,80	**0,83**	**0,84**
	β=6	**0,87**	**0,88**	0,92*	**0,88**	0,91*	0,78	**0,83**	**0,84**
	β=7	**0,87**	**0,88**	0,92*	**0,88**	0,92*	0,77	**0,83**	**0,84**
	β=8	0,86	**0,88**	0,93*	**0,88**	**0,93***	0,75	**0,83**	0,83
	β=9	0,86	**0,88**	0,93*	**0,88**	**0,93***	0,73	0,82	0,82
	β=10	0,85	**0,88**	**0,94***	**0,88**	**0,93***	0,72	0,82	0,81
	β=20	0,81	0,84	0,93*	0,84	0,91*	0,64	0,76	0,75
	β=30	0,79	0,82	0,91*	0,82	0,89	0,62	0,72	0,72
	β=40	0,78	0,80	0,90	0,81	0,87	0,61	0,69	0,71
	β=50	0,78	0,79	0,89	0,80	0,86	0,61	0,68	0,70
	β=60	0,77	0,79	0,88	0,80	0,86	0,61	0,67	0,69
	β=70	0,77	0,78	0,87	0,80	0,85	0,61	0,66	0,69
	β=80	0,77	0,78	0,87	0,80	0,85	0,61	0,65	0,68
	β=90	0,77	0,78	0,87	0,80	0,85	0,61	0,65	0,68
	β=100	0,77	0,78	0,86	0,80	0,85	0,61	0,65	0,68
MW		0,82	0,82	0,87	0,83	0,86	0,74	0,77	0,78
MAX		0,87	0,88	0,94	0,88	0,93	0,82	0,83	0,84
MIN		0,77	0,77	0,81	0,80	0,81	0,61	0,65	0,68

* $p<0,1$

a) First-Choice-Model

b) entspricht dem BTL-Model

Tabelle 5-17: Mittlere absolute Abweichungen (MAA in %) zwischen den beobachteten und prognostizierten Werten der realen Marktanteile bei unterschiedlichen Kaufverhaltensmodellen

	Methode	**ACA**	**CCC-alt**	**CCC-neu**	**CHIC-alt**	**CHIC-neu**	**SE-ACA**	**SE-CCC**	**SE-CHIC**
Kaufverhaltensmodelle	**FC** a)	12,9	14,6	11,4	12,8	11,5	20,3	15,8	17,1
	Logit	6,7	6,6	6,6	6,6	6,6	6,7	6,7	6,7
	β=0,1	6,9	6,9	6,9	6,9	6,9	6,9	6,9	6,9
	β=0,2	6,9	6,9	6,8	6,9	6,8	6,9	6,9	6,9
	β=0,3	6,8	6,8	6,8	6,8	6,8	6,8	6,9	6,8
	β=0,4	6,8	6,7	6,7	6,7	6,7	6,8	6,8	6,8
	β=0,5	6,7	6,7	6,6	6,7	6,6	6,7	6,8	6,7
	β=0,6	6,7	6,6	6,6	6,6	6,6	6,7	6,7	6,7
	β=0,7	6,6	6,6	6,5	6,6	6,5	6,7	6,7	6,6
	β=0,8	6,6	6,6	6,4	6,5	6,5	6,6	6,7	6,6
	β=0,9	6,5	6,5	6,4	6,5	6,4	6,6	6,6	6,5
	β=1 b)	6,5	6,4	6,3	6,5	6,4	6,5	6,6	6,5
	β=2	6,0	5,9	5,7	5,9	5,8	6,0	6,2	6,0
	β=3	5,4	5,3	5,1	5,4	5,2	**5,4**	5,8	5,5
	β=4	4,9	4,8	4,5	4,9	4,6	5,9	5,4	**5,0**
	β=5	**4,8**	**4,4**	4,0	**4,5**	4,0	6,3	**5,0**	5,1
	β=6	5,0	4,6	**3,8**	4,6	**3,9**	6,8	5,2	5,5
	β=7	5,2	4,9	3,9	4,9	4,0	7,2	5,6	6,0
	β=8	5,3	5,2	4,0	5,2	4,1	7,7	5,9	6,3
	β=9	5,5	5,4	4,1	5,4	4,2	8,2	6,3	6,7
	β=10	5,8	5,6	4,2	5,6	4,3	8,7	6,6	7,1
	β=20	7,9	7,0	4,9	7,4	6,1	13,4	8,9	9,3
	β=30	9,9	9,3	6,7	9,7	8,2	16,6	10,2	11,7
	β=40	10,9	10,7	7,8	11,0	9,5	18,2	11,8	13,3
	β=50	11,5	11,6	8,6	11,7	10,2	19,1	13,0	14,2
	β=60	11,9	12,3	9,1	12,2	10,7	19,7	13,8	14,9
	β=70	12,1	12,7	9,5	12,4	11,1	20,0	14,3	15,3
	β=80	12,3	13,0	9,8	12,6	11,3	20,2	14,7	15,7
	β=90	12,4	13,3	10,0	12,7	11,4	20,3	15,0	15,9
	β=100	12,5	13,5	10,2	12,8	11,6	20,4	15,2	16,1
MW		7,9	7,9	6,6	7,8	7,1	10,8	8,6	9,0
MAX		12,9	14,6	11,4	12,8	11,6	20,4	15,8	17,1
MIN		4,8	4,4	3,8	4,5	3,9	5,4	5,0	5,0

a) First-Choice-Model

b) entspricht dem BTL-Model

Vergleicht man jeweils die geringste (Minimum) und die höchste (Maximum) Übereinstimmung zwischen prognostizierten und realen Marktanteilen in Tabelle 5-16 und Tabelle 5-17 wird deutlich, daß auch hier die Wahl der Kaufverhaltensannahme einen entscheidenden Einfluß auf die Güte der Prognose hat. Auffallend ist, daß bei den drei Self-Explicated-Methoden beide Gütemaße wesentlich stärker schwanken als bei den adaptiven Verfahren. Analysiert man die Entwicklung der beiden Gütekriterien, ist festzustellen, daß mit steigendem Exponenten β über den „Optimalwert" die externe Validität aller Verfahren abnimmt bzw. konstant bleibt. Parallel nähern sich die Ergebnisse des exponentiell gewichteten Modells an die Befunde des First-Choice-Model an. Letzteres führt analog zur Prognosevalidität bei allen Methoden zu verhältnismäßig geringer externer Validität. Somit wiederholt sich das schlechte Abschneiden des First-Choice-Model bei der externen Validität.

Das relative Abschneiden der Methoden zueinander wird auch an dieser Stelle bei allen alternativen Kaufverhaltensannahmen überprüft. Die Korrelation der Rangplätze der Methoden ergibt, daß in nur 12% (Pearson) bzw. 3,4% (MAA) aller Fälle unterschiedliche Kaufverhaltensmodelle zum gleichen Ergebnis des Methodenvergleichs führen. Um die Tatsache zu berücksichtigen, daß beide Gütekriterien teilweise nur marginal zwischen alternativen Kaufverhaltensmodellen schwanken, werden die entsprechenden paarweisen Vergleiche zudem anhand des Korrelationskoeffizienten nach Pearson gemessen. Als Grenzwert wird erneut eine Korrelation >0,9 herangezogen. In diesem Fall erhöht sich der Anteil an vergleichbaren Ergebnissen zwischen den Kaufverhaltensannahmen auf 54% (Pearson) bzw. 42% (MAA). Dennoch wird deutlich, daß das Ausmaß der externen Validität und folglich auch der Verfahrensvergleich von der Wahl der Kaufverhaltensannahme beeinflußt wird.

5.4.4 Zusammenfassung der Ergebnisse hinsichtlich der externen Validität

Zur Überprüfung der externen Validität sind die Prognoseergebnisse der Conjoint-Interviews auf die reale Marktsituation übertragen worden. Auf Basis eines Marktanteilsmodells wurden die Marktanteile für 4 verschiedene Marken prognostiziert, wobei über die Kaufwahrscheinlichkeiten für die Marken hinaus verschiedene Marketing-Mix-Instrumente in das Modell eingingen. Als Vergleichs-

größe dienten reale Marktanteilsdaten des GfK-Handelspanel. Wesentliche Ergebnisse der externen Validitätsprüfung sind:

1. Hinsichtlich beider Gütekriterien der externen Validität schneiden die neuen Verfahren CCC-neu und CHIC-neu am besten ab. Nur bei diesen beiden Verfahren ergibt sich eine schwach signifikante Korrelation zwischen beobachteten und prognostizierten Marktanteilen. Die Überlegenheit der beiden neuen Methoden kann folglich zumindest indirekt als (schwach) signifikant interpretiert werden. Alle anderen Methoden weisen eine geringere externe Validität auf, wobei sich bei den drei Self-Explicated-Methoden die geringste Übereinstimmung zwischen prognostizierten und realen Marktanteilen ergibt. Auffallend hierbei ist, daß auch die Unterschiede der Self-Explicated-Methoden zur ACA, CCC-alt und CHIC-alt überraschend gering ausfallen.
2. Der Wahl der Kaufverhaltensannahme kommt bei der Prognose realer Marktanteile eine zentrale Bedeutung zu. Es zeigt sich, daß sowohl die absolute Höhe der externen Validität als auch das relative Abschneiden der Verfahren zueinander und dadurch das Ergebnis des Methodenvergleichs von der Wahl der Kaufverhaltensannahme beeinflußt wird.
3. Im Vergleich zu den Befunden der Prognosevalidität ergibt sich für die Ergebnisse der externen Validität eine wesentlich höhere Konsistenz sowohl zwischen den beiden Operationalisierungen des Validitätsmaßes (Pearson und MAA) als auch hinsichtlich der beiden neuen Berechnungsformeln für die CCC und die CHIC.

Die relativ guten Ergebnisse in ihrer absoluten Höhe dürfen an dieser Stelle allerdings nicht über die grundsätzliche Problematik hinwegtäuschen, die mit der Übertragung der Ergebnisse der Conjoint-Befragung auf die reale Marktsituation verbunden ist. Auch wenn Informationen hinsichtlich verschiedener Marketing-Instrumente in dem Marktanteilsmodell integriert wurden, können nicht alle externen Einflüsse auf Konsumenten in der realen Kaufsituation und auf das gesamte Marktgeschehen berücksichtigt werden. Da sich diese Nichtbeachtung jedoch in gleicher Weise auf alle Methoden auswirkt, werden die Aussagen über das relative Abschneiden der Methoden nur bedingt eingeschränkt und können zumindest in ihrer Tendenz bestehen bleiben.

5.5 Diskussion der empirischen Ergebnisse

Der empirische Validitätsvergleich der untersuchten 8 Methoden wurde anhand verschiedener Gütemaße vorgenommen. Die Ergebnisse hierzu sind in den vorangegangenen Kapiteln jeweils separat diskutiert worden. Für eine abschließende Bewertung der Methoden werden im folgenden alle Validitätsmaße und deren alternative Operationalisierungen gemeinsam betrachtet. Zu diesem Zweck sind alle Ergebnisse in Tabelle 5-18 noch einmal zusammenfassend dargestellt. Bei der Prognosevalidität und der externen Validität ist jeweils das Verfahren hervorgehoben, welches bei entsprechender Operationalisierung die höchste Validität aufweist.

Tabelle 5-18: Gegenüberstellung der Ergebnisse aller betrachteten Gütemaße

Gütemaße und deren Operationalisierungen			**ACA**	**CCC-alt**	**CCC-neu**	**CHIC-alt**	**CHIC-neu**	**SE-ACA**	**SE-CCC**	**SE-CHIC**	**Unterschiede**
Diskriminierende Validität	Bedeutungsgewichte	individuell	• keine Unterschiede zwischen identischen Methoden (SE-CCC und SE-CHIC) • signifikante Unterschiede zwischen allen unterschiedlichen Verfahren								sig.
		aggregiert	• auf aggregierter Ebene (Mittelwerte) nur in Ausnahmefällen Unterschiede zwischen den Verfahren								n.s.
Interne Validität	Conjoint-Phase			0,77 (Kendall)		0,79 (Pearson)		nicht getestet			k.A.
	Kombinationsphase		0,62 (Pearson)	0,68 (Kendall)		0,82 (Pearson)					
Prognosevalidität	individuell	First-Hit-Rate (%)	**45,4**	44,2	41,3	39,3	44,9	40,2	39,7	41,5	n.s.
		Kendalls τ	0,32	**0,33**	**0,33**	0,26	0,27	0,28	0,28	0,29	n.s.
	aggregiert	Pearson	0,64	0,79	**0,89****	0,82*	0,80	0,48	0,79	0,75	bedingt sig.
		MAA (%)	5,9	5,5	6,5	**4,3**	**4,3**	7,2	6,1	4,7	-
Externe Validität	Pearson		0,87	0,88	**0,94***	0,88	0,93*	0,82	0,83	0,84	bedingt sig.
	MAA (%)		4,8	4,4	**3,8**	4,5	3,9	5,4	5,0	5,0	-

sig.: signifikant; n.s.: nicht signifikant; k.A.: keine Aussage möglich; * $p<0,1$; ** $p<0,05$

Faßt man die Befunde der empirischen Validitätsprüfung über alle Kriterien aus Tabelle 5-18 zusammen, können drei zentrale Ergebnisse festgehalten werden:

1. Die beiden neu entwickelten Verfahren CCC und CHIC erweisen sich auf Basis der vorliegenden Untersuchung als geeignete Methoden zur Präferenzmessung. Sie schneiden mit einer nicht signifikanten Ausnahme bei der First-Hit-Rate bei allen betrachteten Gütekriterien gleich gut, im überwiegenden Teil sogar besser ab als das etablierte Verfahren ACA.
2. Als zweites Ergebnis zeigt sich, daß trotz signifikanter Unterschiede zwischen den ermittelten Präferenzstrukturen auf individueller Ebene (vgl. die Ergebnisse zur diskriminierenden Validität), die Unterschiede zwischen den Methoden hinsichtlich ihrer Prognosegüte nur verhältnismäßig gering ausfallen. Während die Unterschiede auf individueller Ebene zwischen allen Verfahren nicht signifikant sind, weisen die beiden neuen Verfahren CCC und CHIC bei aggregierter Analyse eine schwach signifikante Überlegenheit gegenüber der ACA auf, was indirekt über die Korrelationskoeffizienten nach Pearson getestet wurde. Dies gilt sowohl für die Messung der Prognosegüte anhand eines Holdout-Sets als auch bei der Prognose realer Marktdaten.
3. Die unter Punkt 2 angeführten geringen Unterschiede zwischen den Verfahren treten nicht nur im Vergleich der adaptiven hybriden Verfahren untereinander, sondern ebenso gegenüber den untersuchten Self-Explicated-Varianten auf. Damit schneidet die Self-Explicated-Methode in der vorliegenden Untersuchung gegenüber allen drei adaptiven hybriden Methoden gut ab.

Wie bei jeder empirischen Studie sind der Interpretation dieser Ergebnisse Grenzen gesetzt. Hinsichtlich der angewendeten Gütekriterien muß bei der **internen Validität** darauf hingewiesen werden, daß die angegebenen Gütemaße in ihrer absoluten Höhe nicht direkt miteinander vergleichbar sind. Dementsprechend ist ein Test auf Signifikanz der Unterschiede nicht möglich (vgl. letze Spalte in Tabelle 5-18). Für die Self-Explicated-Verfahren ist die interne Validität nicht explizit getestet worden, da die in Tabelle 5-5 ausgewiesenen Werte eher einer Überprüfung der Prognosevalidität und nicht der internen Validität entsprechen. Eine Aussage über das relative Abschneiden der Verfahren kann auf Basis der internen Validität folglich nicht abgeleitet werden. Der Aussagegehalt der Ergebnisse in bezug auf die **Prognosevalidität** wird zumindest teilweise dadurch ein-

geschränkt, daß die Prognose der simulierten Kaufentscheidung insgesamt nur unzureichend gelingt. Die sehr niedrigen Werte für die First-Hit-Rate und die durchschnittlichen Kendalls τ-Werte auf individueller Ebene bei allen Methoden deuten darauf hin, daß starke Reliabilitätsprobleme in den Antworten der Probanden oder gar Fehler in der Datenerhebung zu vermuten sind. Weiterhin muß berücksichtigt werden, daß, obgleich eine verhältnismäßig große Stichprobe befragt wurde, vor allem hinsichtlich der **externen Validität** nicht ohne weiteres sichergestellt werden kann, daß diese repräsentativ für einen Vergleich mit den realen Marktdaten des GfK-Handelspanel ist. Ferner wurde in dieser Untersuchung lediglich eine **Produktart** mit einer bestimmten **Anzahl an Eigenschaften** betrachtet. Wenn auch in dem vorgenommenen Literaturüberblick über bisherige Vergleichsstudien in Kapitel 2 keine systematischen Einflüsse dieser Merkmale einer Studie aufgezeigt werden konnten, kann nicht davon ausgegangenen werden, daß die vorliegenden Ergebnisse auf andere Untersuchungsgegenstände ohne weiteres übertragbar sind.

Trotz dieser Einschränkungen ermutigen die vorliegenden Befunde, die beiden neuen Verfahren im Rahmen weiterer Untersuchungen dahingehend zu überprüfen, ob sie für alternative Fragestellungen ebenso zufriedenstellende Ergebnisse liefern. Für weitere Untersuchungen sei gleichzeitig empfohlen, trotz des zusätzlichen Erhebungsaufwands eine explizite Überprüfung der Reliabilität in den Interviewablauf zu integrieren. Dann können Verzerrungen der Prognoseergebnisse aufgrund fehlender Reliabilität identifiziert und die beiden Effekte bei der Ergebnisinterpretation getrennt werden.

6 Fazit der Arbeit

Während kein Zweifel darüber besteht, daß Kenntnisse über die Präferenzstrukturen der Nachfrager von zentraler Bedeutung für ein erfolgreiches Produkt-Management sind, gibt es bisher kein Verfahren, das für deren Messung eindeutig empfohlen werden kann. Der vorgenommene Literaturüberblick über empirische Validitätsvergleiche führte zu dem Ergebnis, daß sich keine Variante der Conjoint-Analyse, der Self-Explicated-Methode oder der hybriden Verfahren als eindeutig überlegen erwies. Insbesondere für die Adaptive Conjoint Analysis (ACA), die sich mittlerweile als Standardverfahren zur Präferenzmessung etabliert hat, konnte gezeigt werden, daß sie in zahlreichen Studien schlechter als die verglichenen Methoden abschneidet.

Vor diesem Hintergrund war ein Ziel der vorliegenden Arbeit, neue computergestützte Conjoint-Varianten zu entwickeln. Motivation für die Entwicklung der neuen Verfahren waren vornehmlich zwei Gründe. Erstens wirft das relativ schlechte Abschneiden der ACA bei empirischen Vergleichen sowohl gegenüber der rein dekompositionellen Conjoint-Analyse als auch der Self-Explicated-Methode vor allem die Frage auf, ob methodische Schwächen des Verfahrens dafür verantwortlich sind. Zweitens wird mit den neuen Methoden der Vielfalt dekompositioneller Varianten auch im Bereich computergestützter hybrider Methoden Rechnung getragen, wobei neuere Entwicklungen wie die Choice-Based Conjoint Analysis integriert wurden. Mit der Customized Computerized Conjoint Analysis (CCC) und der Choice-Oriented Individualized Conjoint Analysis (CHIC) wurden zwei computergestützte hybride Verfahren zur Präferenzmessung vorgestellt, die sich wie die ACA durch einen individualisierten Interviewablauf auszeichnen. Sie sind dadurch auch für Anwendungen mit vielen Eigenschaften geeignet und werden so einer wichtigen Anforderung der Praxis gerecht.

Über die Entwicklung der neuen Methoden hinaus war ein zweites Ziel der Arbeit, die praktische Anwendbarkeit der neuen Verfahren zu überprüfen. In einer umfassenden empirischen Studie mit einer Stichprobe von 748 Probanden wurde eine vergleichende Validitätsprüfung vorgenommen. Als Referenzgrößen für die Ergebnisgüte der CCC und der CHIC wurden zum einen die ACA als am häufig-

sten eingesetztes Verfahren und zum anderen die Self-Explicated-Methode herangezogen. Letztgenannter konnte in bisherigen empirischen Vergleichsstudien eine vergleichbare Prognosegüte wie den Conjoint-Varianten bescheinigt werden.

Die empirische Überprüfung der untersuchten Methoden erfolgte anhand verschiedener Gütekriterien. In einem ersten Schritt wurde mit der diskriminierenden Validität getestet, ob die verschiedenen Methoden zur Präferenzmessung zu unterschiedlichen Ergebnissen in den geschätzten Präferenzstrukturen führen. Da dies der Fall war, wurde mit der internen Validität die Modellgüte der Verfahren an sich überprüft und auf Basis einer simulierten Kaufentscheidung die Prognosevalidität der Verfahren verglichen. Eine Besonderheit der vorliegenden Untersuchung liegt in der Überprüfung der externen Validität durch Prognose der realen Marktanteile für die untersuchten Marken.

Als ein zentrales Ergebnis der empirischen Studie kann festgehalten werden, daß die beiden neuen Verfahren CCC und CHIC hinsichtlich der angewendeten Gütemaße zu wenigstens gleich guten, teilweise sogar besseren Prognoseergebnissen führen als das etablierte Verfahren ACA. Betrachtet man das in der Praxis am häufigsten eingesetzte Verfahren als Referenzgröße, erweisen sich die neu entwickelten Verfahren als geeignete Methoden zur Präferenzmessung. Die teilweise geringen Unterschiede in den Ergebnissen geben aber noch keine klare Antwort auf die Frage, ob die Schwachstellen der ACA durch die alternativen Vorgehensweisen der CCC und der CHIC behoben werden. Die vorliegenden Befunde deuten zwar tendenziell darauf hin, aber erst weitere Studien müssen zeigen, ob sich die Ergebnisse stabilisieren.

Ein zweites, vor allem unter Anwendungsgesichtspunkten wichtiges Ergebnis resultiert aus den guten empirischen Befunden für die Self-Explicated-Methode. Die jeweils separate Auswertung der Daten nach der ersten Interviewphase, dem kompositionellen Befragungsteil, führte in der vorliegenden Studie gegenüber allen drei adaptiven Verfahren zu vergleichbaren Prognoseergebnissen. Diese Befunde sind vor allem für die praktische Anwendung in der Marktforschung von besonderem Interesse. Die Implementierung der Self-Explicated-Methode zur Präferenzmessung ist wesentlich einfacher als ein Conjoint-Interview. Ein damit verbundener Vorteil des Verfahrens besteht in der Einsatzmöglichkeit von Tele-

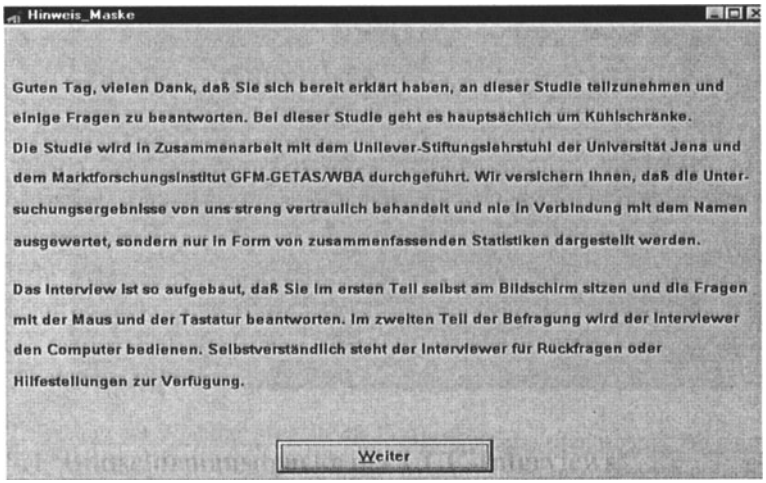
foninterviews, was im Rahmen der Conjoint-Analyse immer wieder gewünscht wird, in der Praxis jedoch nur sehr eingeschränkt durchführbar ist.

Zusammenfassend muß angesichts dieser Ergebnisse die eindeutige Dominanz der ACA in der praktischen Anwendung kritisch hinterfragt werden. Die vorliegende Arbeit liefert zum einen neue adaptive Verfahren, die genauso gut, teilweise sogar besser geeignet sind, die Präferenzstrukturen von Nachfragern zu messen. Zum anderen stellt die Self-Explicated-Methode ein alternatives Verfahren zur Präferenzmessung dar, das aufgrund von Zeit- und Kostenvorteilen unter Anwendungsgesichtspunkten allen Conjoint-Varianten überlegen ist.

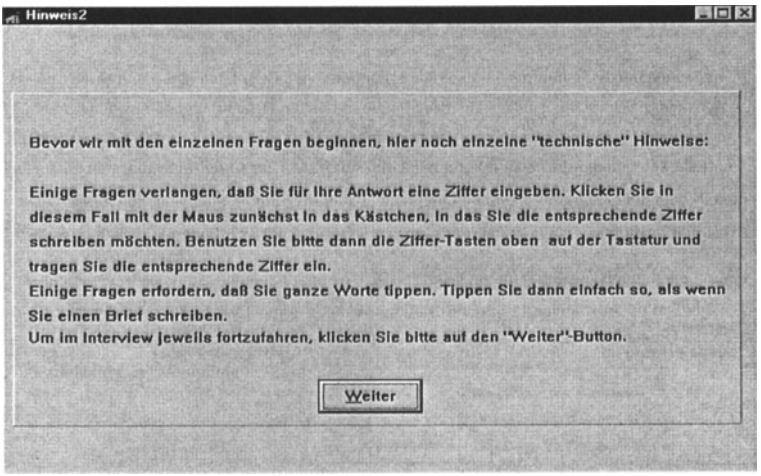
7 Anhang

7.1 Bildschirmausdrucke des CCC-Interviews[360]

Begrüßungsbildschirm



Hinweisbildschirm zur Bedienung des Programms



---

360 Es sei darauf hingewiesen, daß die einzelnen Befragungsschritte bei allen drei Interviews jeweils exemplarisch anhand eines Beispiels und nicht für alle Eigenschaften bzw. alle Eigenschaftskombinationen in der Conjoint- bzw. Kombinationsphase dargestellt werden.

Self-Explicated-Phase

Hinweisbildschirm

Info Maske 1

Wir möchten gerne von Ihnen erfahren, wie wichtig die Eigenschaften Marke, Energieverbrauch, Sterne des Gefrierfachs, Preis, Nutzinhalt, Abtaumöglichkeit, Garantie und Innenausstattung für Sie beim Kauf eines Kühlschrankes sind.

Im folgenden zeigen wir Ihnen dazu diese Eigenschaften und deren Ausprägungen, die Sie jeweils beurteilen sollen.

Gehen Sie bitte im Verlauf des gesamten Interviews immer davon aus, daß es sich um einen FCKW- und FKW-freien Kühlschrank mit Gefrierfach handelt.
Des weiteren weist jeder Kühlschrank die Außenmaße Höhe/Breite/Tiefe 80cm/60cm/60cm auf und ist ein- bzw. unterbaufähig.

Weiter

<u>Bewertung der Eigenschaftsausprägungen:</u>
Identifikation der besten und schlechtesten Ausprägung

Dieser Befragungsschritt entfiel für die quantitativen Eigenschaften Preis, Energieverbrauch, Garantie und Nutzinhalt.

Daten-Erfassung Maske 1

Welche der folgenden Ausprägungen bei der Eigenschaft: Marke würde Ihnen am besten gefallen und welche würde Ihnen am wenigsten gefallen, sofern alle anderen Eigenschaften gleich sind?

Schreiben Sie bitte jeweils die entsprechende Nummer der zutreffenden Ausprägung in das dafür vorgesehene Kästchen.

Nr.	Ausprägungen
1	Liebherr
2	Siemens
3	Electrolux
4	Quelle Privileg

Am besten würde mir gefallen Nr. ...

Am wenigsten würde mir gefallen Nr. ...

Weiter

Bewertung der Eigenschaftsausprägungen:

Bewertung der mittleren Ausprägung anhand einer 11-Punkte-Rating-Skala bei der die beste und die schlechteste Ausprägung als Endpole der Skala ausgewiesen sind.

Daten-Erfassung Maske 2

Sie sehen hier eine Skala von 10 bis 0. Dabei wurde der 10 die Ausprägung zugeordnet, die Ihnen am besten gefallen würde. Der Ausprägung, die Ihnen am wenigsten gefallen würde, ist der Wert 0 zugeordnet worden. Auch hier gilt jeweils, daß alle anderen Eigenschaften gleich sind.

Bitte beurteilen Sie jetzt die übrigen Ausprägungen, indem Sie Ihnen einen Wert auf der Skala zuordnen. Je höher der Wert ist, desto besser gefällt Ihnen diese Ausprägung.

Dabei darf jeder Wert nur einmal vergeben werden!

Bitte tragen Sie den Skalenwert neben die Ausprägung in das dafür vorgesehende Kästchen in der untenstehenden Tabelle ein.

10 — 9 — 8 — 7 — 6 — 5 — 4 — 3 — 2 — 1 — 0

gefällt mir am besten — gefällt mir am wenigsten

Quelle Privileg — Liebherr

Nr.	Auspraegungen	Beurteilungswert
1	Siemens	
2	Electrolux	

Eigenschaft: Marke

Weiter

Erfassung der Wichtigkeiten für die Eigenschaften:

1. Frage nach der wichtigsten Eigenschaft (10 Punkte)

2. Vergabe eines Punktwertes von 0 bis 10 Punkte für die übrigen Eigenschaften unter Verwendung der wichtigsten Eigenschaft als Ankergröße

Daten-Erfassung Maske 4

Sie sehen in der untenstehenden Tabelle alle Eigenschaften mit der jeweils für Sie besten und schlechtesten Ausprägung. Wir möchten gerne von Ihnen erfahren, wie wichtig diese Eigenschaften für Sie sind.

Bitte wählen Sie hierfür zunächst die Eigenschaft aus, die für Sie am wichtigsten ist und ordnen dieser Eigenschaft den Wert 10 zu. Die wichtigste Eigenschaft sollte also diejenige sein, bei der Ihnen die Verbesserung von der schlechtesten Ausprägung zur besten Ausprägung am wertvollsten ist.

Bitte beurteilen Sie anschließend die anderen Eigenschaften im Vergleich dazu. Ordnen Sie ihnen bitte einen Wert zwischen 0 und 10 zu. Je höher der Wert ist, den Sie zuordnen, desto wichtiger ist ihnen die Eigenschaft.

Sofern Ihnen verschiedene Eigenschaften gleich wichtig sind, können Sie diesen die gleichen Werte zuteilen. Tragen Sie die Werte bitte in die letzte Spalte der Tabelle ein.

Nr.	Eigenschaften	schlechteste Ausprägung	beste Ausprägung	Wichtigkeit
1	Marke	Liebherr	Quelle Privileg	
2	Energieverbrauch	250 kWh pro Jahr	150 kWh pro Jahr	
3	Sterne des Gefrierfachs	*-Fach (-6 °C)	***-Fach (-18 °C)	
4	Preis	649,- DM	449,- DM	
5	Nutzinhalt	115 Liter Nutzinhalt	145 Liter Nutzinhalt	
6	Abtaumöglichkeit	Abtauung von Hand	Abtau-Automatik im Kühlteil	
7	Garantie	12 Monate Garantie	24 Monate Garantie	
8	Innenausstattung	Innenaufteilung vorgegeben	Kühlteil und Tür frei gestaltb	

Weiter

Conjoint-Phase

Hinweisbildschirm mit Erläuterung der Conjoint-Aufgabe

Info-Maske 2

Sie sehen gleich 9 verschiedene Karten, die unterschiedliche Kühlschränke beschreiben.
Gehen Sie bitte hierbei wieder davon aus, daß es sich immer um einen FCKW- und FKW-freien Kühlschrank mit Gefrierfach handelt. Des weiteren weist jeder Kühlschrank die Außenmaße Höhe/Breite/Tiefe 80cm/60cm/60cm auf und ist ein- bzw. unterbaufähig.

Bringen Sie die 9 Karten bitte in die Reihenfolge, in der Sie die Kühlschränke kaufen würden. Das bedeutet, der Rang 1 wird dem Kühlschrank (bzw. der Karte) zugeordnet, den Sie als erstes kaufen würden.
Der Rang 2 wird dem Kühlschrank zugeordnet, den Sie an zweiter Stelle kaufen würden, u.s.w..

Nachdem Sie die Ränge für die einzelnen Karten eingegeben haben, erscheinen die 9 Karten in der Reihenfolge, in der Sie die Kühlschränke am ehesten kaufen würden. Geben Sie dann bitte an, wie wahrscheinlich es wäre, daß Sie diesen Kühlschrank tatsächlich kaufen würden, sofern Sie vor der Kaufentscheidung stünden und dieser derzeit erhältlich wäre.

Weiter

Erfassung ganzheitlicher Präferenzen (1. Schritt):
Angabe einer Präferenzrangfolge für 9 Produktalternativen, die durch 4 individuell verschiedene Kerneigenschaften beschrieben sind.

Kartenliste

250 kWh pro Jahr	250 kWh pro Jahr	250 kWh pro Jahr
Liebherr	Electrolux	Quelle Privileg
649,- DM	549,- DM	449,- DM
*-Fach (-6 °C)	*-Fach (-6 °C)	***-Fach (-18 °C)
Karte4	**Karte5**	**Karte6**
200 kWh pro Jahr	200 kWh pro Jahr	200 kWh pro Jahr
Liebherr	Electrolux	Quelle Privileg
549,- DM	449,- DM	649,- DM
***-Fach (-18 °C)	*-Fach (-6 °C)	*-Fach (-6 °C)
Karte7	**Karte8**	**Karte9**
150 kWh pro Jahr	150 kWh pro Jahr	150 kWh pro Jahr
Liebherr	Electrolux	Quelle Privileg
449,- DM	649,- DM	549,- DM
*-Fach (-6 °C)	***-Fach (-18 °C)	*-Fach (-6 °C)

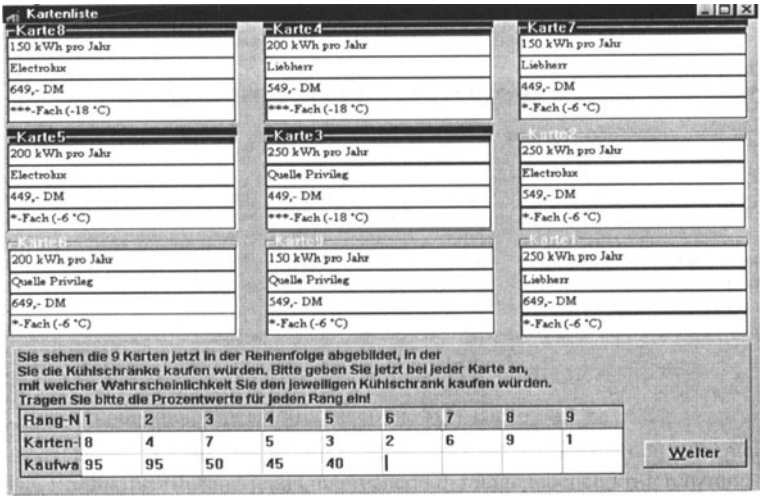
Geben Sie bitte die Reihenfolge ein, in der Sie die Kühlschränke kaufen würden!

Rang-N	1	2	3	4	5	6	7	8	9
Karten-	8	4	7						

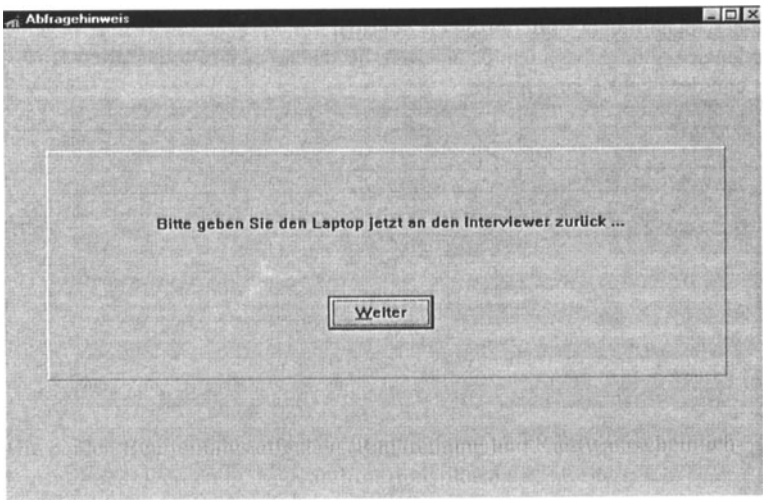
Weiter

Erfassung ganzheitlicher Präferenzen (2. Schritt):

Angabe einer Kaufwahrscheinlichkeit von 0% bis 100% für jeden Rangplatz. Die Produktprofile sind jetzt entsprechend der angegebenen Präferenzrangfolge auf dem Bildschirm sortiert.



Die beiden Beurteilungsaufgaben (Rangreihung und Kaufwahrscheinlichkeitsabfrage) werden in der **Kombinationsphase** für 9 weitere Produktprofile wiederholt.

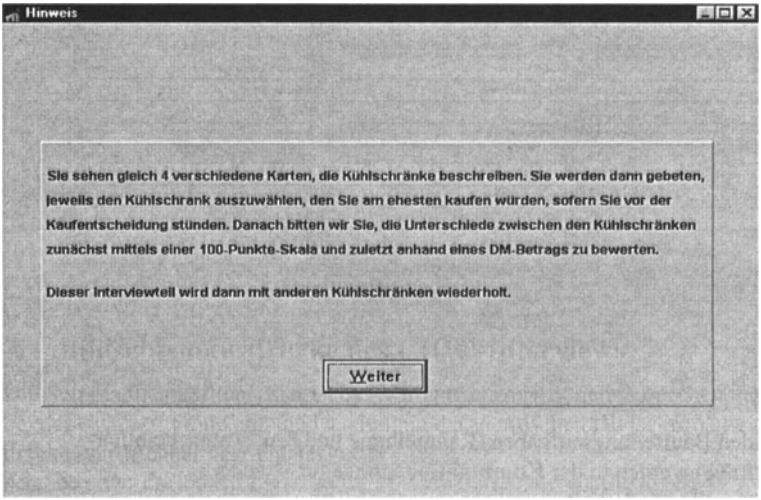


Alle weiteren erhobenen Zusatzdaten hinsichtlich der Probanden wurden bei allen drei Methoden (CCC, CHIC und ACA) von den Interviewern direkt eingegeben. Die entsprechenden Bildschirmausdrucke werden aus diesem Grund nicht mehr abgebildet.

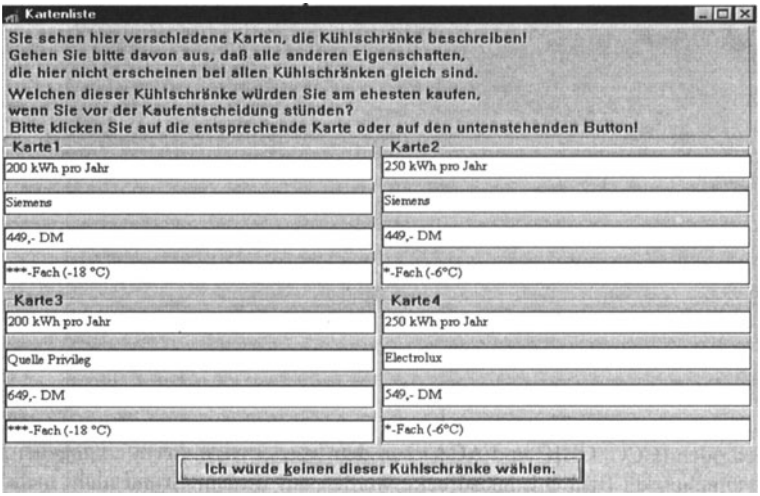
7.2 Bildschirmausdrucke des CHIC-Interviews

Die Self-Explicated-Phase der CHIC-Variante entspricht dem Vorgehen bei der CCC. Auf eine erneute Darstellung der identischen Bildschirmausdrucke wurde daher verzichtet.

Conjoint-Phase



<u>Erfassung ganzheitlicher Präferenzen (1. Schritt):</u>
Auswahlentscheidung aus einem Choice-Set, bestehend aus 4 Produktalternativen und der Nicht-Kauf-Option.



Wiederholte Auswahlentscheidung aus den verbliebenen 3 Produkt-alternativen

Kartenliste

Angenommen, Ihnen stehen nur noch diese 3 Kühlschränke zur Wahl. Welchen dieser 3 Kühlschränke würden Sie dann am ehesten kaufen?
Bitte klicken Sie auf die entsprechende Karte!

Karte2	Karte3	Karte4
250 kWh pro Jahr	200 kWh pro Jahr	250 kWh pro Jahr
Siemens	Quelle Privileg	Electrolux
449,- DM	649,- DM	549,- DM
*-Fach (-6°C)	***-Fach (-18 °C)	*-Fach (-6°C)

Entsprechende Auswahlentscheidung aus den letzten beiden Produkt-alternativen

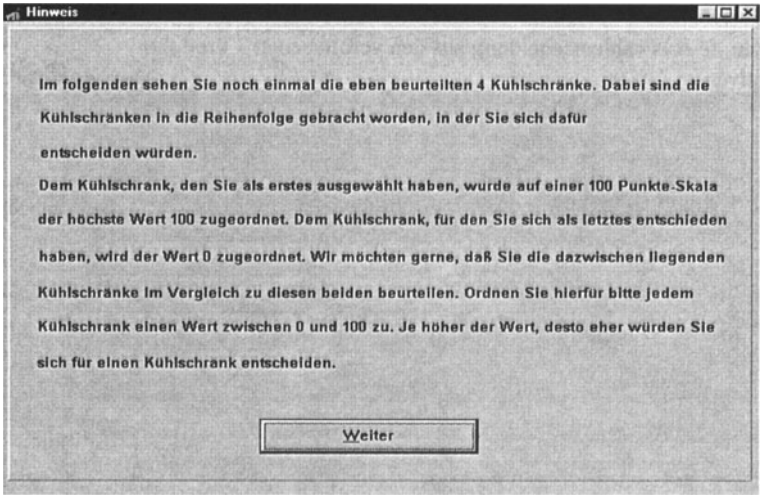
Kartenliste

Wiederum angenommen, Ihnen stehen nur noch diese 2 Kühlschränke zur Auswahl. Welchen dieser 2 Kühlschränke würden Sie dann am ehesten kaufen?

Bitte klicken Sie auf die entsprechende Karte!

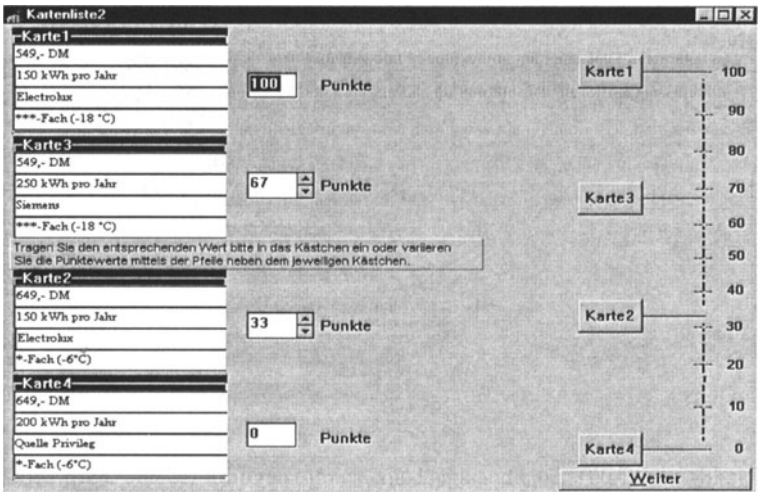
Karte2	Karte4
250 kWh pro Jahr	250 kWh pro Jahr
Siemens	Electrolux
449,- DM	549,- DM
*-Fach (-6°C)	*-Fach (-6°C)

Hinweisbildschirm mit Erläuterung der folgenden Aufgabe

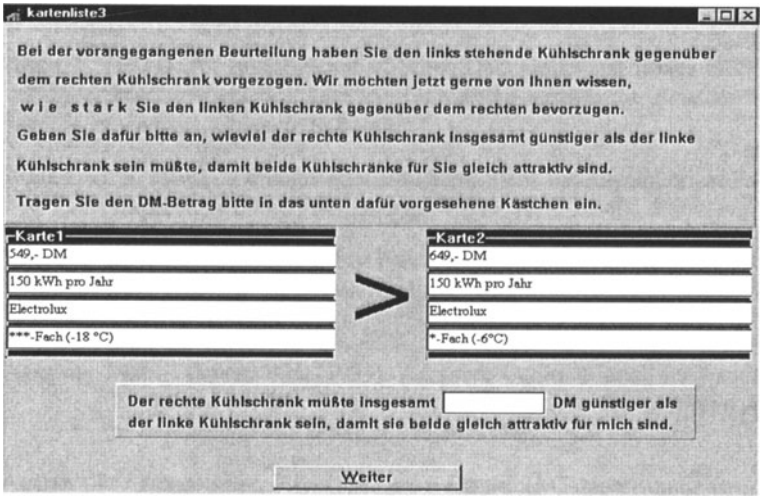


2. Schritt:

Quantifizierung der Präferenz-unterschiede anhand einer 100-Punkte-Skala
Die Produktprofile sind jetzt in der ausgewählten Reihenfolge abgebildet



3. Schritt:
Transformation der Präferenz-unterschiede in eine monetäre Größe



Diese 3 Befragungsschritte werden für jedes zu bewertende Choice-Set der Conjoint-Phase sowie für ein zusätzliches für die **Kombinationsphase** wiederholt.

Die Anzahl der insgesamt zu bewertenden Choice-Sets wird von dem Programm in Abhängigkeit der individuellen Kerneigenschaften und deren Ausprägungen festgelegt.

Nach Bewertung aller Choice-Sets werden die Probanden entsprechend dem Vorgehen bei der CCC gebeten, den Laptop wieder an den Interviewer zu übergeben, der alle weiteren Zusatzdaten hinsichtlich der befragten Person direkt eingibt.

Literaturverzeichnis

Acito, F. (1977): An Investigation of Some Data Collection Issues in Conjoint Measurement, in: *American Marketing Association Educators' Proceedings*, Chicago, S. 82-85.

Addelman, S. (1962): Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments, in: *Technometrics*, 4 (No. 1), S. 21-46.

Agarwal, M.K. (1989): How many Pairs Should we use in Adaptive Conjoint Analysis: An Empirical Analysis, Winter AMA Conference, St. Petersburg, Florida, S. 1-10.

Agarwal, M.K. / Green, P.E. (1991): Adaptive Conjoint Analysis versus Self-Explicated Models: Some Empirical Results, in: *International Journal of Research in Marketing* (8), S. 141-146.

Akaah, I.P. / Korgaonkar, P.K. (1983): An Empirical Comparison of the Predictive Validity of Self-Explicated, Huber-Hybrid, Traditional Conjoint, and Hybrid Conjoint Models, in: *Journal of Marketing Research*, 20 (May), S. 187-197.

Albers, S. / Brockhoff, K. (1985): Die Gültigkeit der Ergebnisse eines Testmarktsimulators bei unterschiedlichen Daten und Auswertungsmethoden, in: *Zeitschrift für betriebswirtschaftliche Forschung*, 37, S. 191-217.

Baalbaki, I.B. / Malhotra, N.K. (1995): Standardization versus Customization in International Marketing: An Investigation Using Bridging Conjoint Analysis, in: *Journal of the Academy of Marketing Science*, 23 (3), S. 182-194.

Backhaus, K. / Erichson, B. / Plinke, W. / Weiber, R. (1996): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*, 8. Aufl., Berlin et al.

Baier, D. / Säuberlich, F. (1997): Kundennutzenschätzung mittels individueller Hybrid-Conjointanalyse, in: *Zeitschrift für betriebswirtschaftliche Forschung*, 49, S. 951-972.

Bateson, J.E.G. / Reibstein, D. / Boulding, W. (1987): Conjoint Analysis Reliability and Validity: A Framework for Future Research, in: *Review of Marketing*, M. J. Houston, ed. Chicago, American Marketing Association, S. 451-481.

Batsell, R.R. / Louviere, J.J. (1991): Experimental Analysis of Choice, in: *Marketing Letters*, 2:3, S. 199-214.

Bleicker, U. (1983): *Produktbeurteilung der Konsumenten*, Würzburg, Wien.

Böcker, F. (1986): Präferenzforschung als Mittel marktorientierter Unternehmensführung, in: *Zeitschrift für betriebswirtschaftliche Forschung*, 38, S. 543-574.

Boecker, F. / Schweikl, H. (1988): Better Preference Prediction with Individualized Sets of Relevant Attributes, in: *International Journal of Research in Marketing*, 5, S. 15-24.

Böhler, H. (1992): *Marktforschung*, 2. überarb. Aufl., Stuttgart.

Bond, J.R.P. (1991): Increasing the Value of Computer Interviewing, ESOMAR Congress Marketing in the Europe, Luxembourg, S. 737-754.

Bortz, J. (1993): *Statistik für Sozialwissenschaftler*, 4., vollst. überarb. Aufl., Berlin et al.

Bretton-Clark (1989): CONJOINT- LINMAP, Morristown, NJ.

Brockhoff, K. (1999): *Produktpolitik*, 4., vollst. überarb. Aufl., Stuttgart.

Carmone, F. (1987), ACA System for Adaptive Conjoint Analysis, in: *Journal of Marketing Research*, 8 (August) 24, S. 325-327.

Carmone, F. / Schaffer, C.M. (1995): Software Review: Conjoint Designer Version 3, Semigraf Version 2, Conjoint Linmap, Conjoint Segmenter, Bridger, Consurv, Adaptive Conjoint Analysis System, Choice Based Conjoint System, Conjoint Value Analysis System, in: *Journal of Marketing Research*, 32 (February), S. 113-120.

Carmone, F.J. / Green P.E. / Jain, A.K. (1978): Robustness of Conjoint Analysis: Some Monte Carlo Results, in: *Journal of Marketing Research*, 15 (May), S. 300-303.

Cattin, P. / Hermet, G. / Pioche, A. (1982): Alternative Hybrid Models for Conjoint Analysis: Some Empirical Results, in: *Analytical Approaches to Product and Market Planning: The Second Conference*, Cambridge, MA: Marketing Science Institute (October), S. 142-152.

Chrzan, K. (1994): Three Kinds of Order Effects in Choice-Based Conjoint Analysis, in: *Marketing Letters*, 5 (2), S. 165-172.

Clement, M. (2000): *Analyse und Prognose der Nutzung des Interaktiven Fernsehens*, Wiesbaden.

Curry, J. (1998): Understanding Conjoint Analysis, URL: http://www.sawtooth.com/news/library/articles/predict.html, Sawtooth Software.

DeSarbo, W.S. / Ramaswamy, V. / Cohen, S.H. (1995): Market Segmentation with Choice-Based Conjoint Analysis, in: *Marketing Letters*, 6:2, S. 137-147.

Dorsch, M. / Teas, R.K. (1992): A Test of the Convergent Validity of Self-Explicated and Decompositional Conjoint Measurement, in: *Journal of the Academy of Marketing Science*, 20 (1), S. 37-48.

Eisenführ, F. / Weber, M. (1994): *Rationales Entscheiden*, 2. Aufl., Berlin et al.

Elrod, T. / Kumar, S.K. (1989): Bias in the First Choice Rule for Predicting Share, in: *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, ed. Ketchum, ID: Sawtooth Software, Inc., S. 259-271.

Elrod, T. / Louviere, J.J. / Davey, K.S. (1992): An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models, in: *Journal of Marketing Research*, 29 (August), S. 368-377.

Erichson, B. (1979): Prognose für neue Produkte, in: *Marketing Zeitschrift für Forschung und Praxis*, Heft 4, S. 255-266.

Ernst, O. / Sattler, H. (2000): Multimediale versus traditionelle Conjoint Analysen: Ein empirischer Vergleich alternativer Präsentationsformen, erscheint in: *Marketing Zeitschrift für Forschung & Praxis*, 70. Jg.

Falk, M. / Becker, R. / Marohn, F. (1995): *Angewandte Statistik mit SAS. Eine Einführung*, Berlin et al.

Finkbeiner, C.T. (1988): Comparison of Conjoint Choice Simulators, in: *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, ed. Ketchum, ID: Sawtooth Software, Inc., S. 75-108.

Fischer, G.W. (1995): Range Sensitivity of Attributes Weights in Multiattribute Value Models, in: *Organizational Behavior and Human Decision Progress*, 63, No.3, S. 252-266.

Gaul, W. / Baier D. / Aperis A. (1996): Verfahren der Testmarktsimulation in Deutschland: Eine vergleichende Analyse, in: *Marketing Zeitschrift für Forschung und Praxis*, Heft 3, S. 203-217.

Gedenk, K. / Sattler, H. (1999): The Impact of Price Thresholds on Profit Contribution - Are 9-Ending Prices Based on Tradition or Economic Rationale?, in: *Journal of Retailing* (JR), 75 (1), S. 33-57.

Green, P.E. (1984): Hybrid Models for Conjoint Analysis: An Expository Review, in: *Journal of Marketing Research*, 21 (May) , S. 155-169.

Green, P.E. / Carmone, F.J. / Wind, Y. (1972): Subjective Evaluation Models and Conjoint Measurement, in: *Behavioral Science*, 17, S. 288-299.

Green, P.E. / Goldberg S.M. / Montemayer, M. (1981): A Hybrid Utility Estimation Model for Conjoint Analysis, in: *Journal of Marketing*, 45, S. 33-41.

Green, P.E. / Goldberg, S.M. / Wiley, J.B. (1982): A Cross-Validation Test of Hybrid Conjoint Models, in: Proceedings of 1982 Annual Meeting of the Association for Consumer Research, (October), San Francisco, S. 147-150.

Green, P.E. / Helsen, K. (1989): Cross-Validation Assessment of Alternatives to Individual-Level Conjoint Analysis: A Case Study, in: *Journal of Marketing Research*, 26 (August), S. 346-350.

Green, P.E. / Krieger, A.M. (1988): Choice Rules and Sensitivity Analysis in Conjoint Simulators, in: *Journal of the Academy of Marketing Science*, 16 (1), S. 114-127.

Green, P.E. / Krieger, A.M. (1996): Individualized Hybrid Models for Conjoint Analysis, in: *Management Science*, 42 (June), S. 850-867.

Green, P.E. / Krieger A.M. / Agarwal, M. (1991): Adaptive Conjoint Analysis: Some Caveats and Suggestions, in: *Journal of Marketing Research*, 28, S. 215-222.

Green, P.E. / Krieger, A.M. / Agarwal, M.K. (1993): A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences, in: *Marketing Letters*, 4:4, S. 369-380.

Green, P.E. / Krieger, A.M. / Bansal, P. (1988): Completely Unacceptable Levels in Conjoint Analysis: A Cautionary Note, in: *Journal of Marketing Research*, 25 (August), S. 293-300.

Green, P.E. / Schaffer, C.M. / Patterson, K.M. (1991): A Validation Study of Sawtooth Softwares Adaptive Conjoint Analysis (ACA), in: *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, ed. Ketchum, ID: Sawtooth Software, Inc., S. 303-314.

Green, P.E. / Srinivasan, V. (1978): Conjoint Analysis in Consumer Research: Issues and Outlook, in: *Journal of Consumer Research*, 5, S. 103-123.

Green, P.E. / Srinivasan, V. (1990): Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice, in: *Journal of Marketing*, 54, S. 3-19.

Gustafsson, A. / Herrmann, A. / Huber, F. (2000): Conjoint Analysis as an Instrument of Market Research Practice, in: A. Gustafsson, A. Herrmann und F. Huber (Hrsg.): *Conjoint Measurement: Methods and Applications*, Berlin et al., S. 4-45.

Gutsche, J. (1995): *Produktpräferenzanalyse: Ein modelltheoretisches und methodisches Konzept zur Marktsimulation mittels Präferenzerfassungsmodellen*, Berlin.

Hahn, C. (1997): *Conjoint- und Discrete Choice-Analyse als Verfahren zur Abbildung von Präferenzstrukturen und Produktwahlentscheidungen: Ein theoretischer und computergestützter empirischer Verg*leich, Münster.

Hammann P. / Erichson B. (1994): *Marktforschung*, 3. überarb. und erw. Aufl., Stuttgart et al.

Hartung, J. (1991): *Statistik - Lehr- und Handbuch der angewandten Statistik*, 8., durchges. Aufl., München et al.

Heeler, R.M. / Okechuku, C. / Reid, S. (1979): Attribute Importance: Contrasting Measurement, in: *Journal of Marketing Research*, 16 (February), S. 60-63.

Hensel-Börner, S. / Sattler H. (2000): Ein empirischer Validitätsvergleich zwischen der Customized Computerized Conjoint Analysis (CCC), der Adaptive Conjoint Analysis (ACA) und Self-Explicated-Verfahren; *Zeitschrift für Betriebswirtschaft*, 70. Jg. (6), S. 705-727.

Herrmann, A. / Schmidt-Gallas, D. / Huber, F. (2000): Adaptive Conjoint Analysis: Understanding the Methodology and Assessing Reliability and Validity, in: A. Gustafsson, A. Herrmann und F. Huber (Hrsg.): *Conjoint Measurement: Methods and Applications*, Berlin et al., S. 253-277.

Hoepfl, R.T. / Huber, G.P. (1970): A Study of Self-Explicated Utility Models, in: *Behavioral Science*, 15, S. 408-414.

Huber, F. / Herrmann, A. / Gustafsson, A. (2000): On the Influence of the Evaluation Methods in Conjoint Design - Some Empirical Results, in: A. Gustafsson, A. Herrmann und F. Huber (Hrsg.): *Conjoint Measurement: Methods and Applications*, Berlin et al., S. 183-208.

Huber, G.P. / Daneshgar, R. / Ford, D.L. (1971): An Empirical Comparison of Five Utility Models for Predicting Job Preferences, in: *Organizational Behavior and Human Performance*, 6, S. 267-282.

Huber, J. / Wittink, D.R. / Fiedler, J.A. / Miller, R. (1993): The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice, in: *Journal of Marketing Research*, 30 (February), S. 105-114.

Huber, J. / Zwerina, K. (1996): The Importance of Utility Balance in Efficient Choice Designs, in: *Journal of Marketing Research*, 33, S. 307-317.

Jacoby, J. / Olson, J.C. / Haddock, R.A. (1971): Price, Brand Name and Product Composition Characteristics as Determinants of Perceived Quality, in: *Journal of Applied Psychology*, 55, S. 570-579.

Jain, A.K. / Acito, F. / Malhotra, N.K. / Mahajan, V. (1979): A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models, in: *Journal of Marketing Research*, 16 (August), S. 313-322.

Jain, A.K. / Mahajan, V. / Malhotra, N.K. (1979): Multiattribute Preference Models for Consumer Research: A Synthesis, in: *Conference Proceedings, Association for Consumer Research*, S. 248-252.

Johnson, R.M. (1974): Trade-off Analysis of Consumer Values, in: *Journal of Marketing Research*, 11 (May), S. 121-127.

Johnson, R.M. (1987): Adaptive Conjoint Analysis, in: *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, Ketchum, ID: Sawtooth Software, Inc., S. 253-265.

Johnson, R.M. (1989): Assessing the Validity of Conjoint Analysis, in: *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, Ketchum, ID: Sawtooth Software, Inc., S. 273-280.

Johnson, R.M. (1991): Comment on Adaptive Conjoint Analysis: Some Caveats and Suggestions, in: *Journal of Marketing Research*, 28, S. 223-225.

Johnson, R.M. (1997a): Including Holdout Choice Task in Conjoint Studies, Sawtooth Software, Technical Paper.

Johnson, R.M. (1997b): Individual Utilities from Choice Data: A New Method, in: *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, ed. Ketchum, ID: Sawtooth Software, Inc., S. 191-208.

Johnson, R.M. / Orme, B.K. (1996): How Many Questions Should You Ask In Choice-Based Conjoint Studies?, Sawtooth Software, Technical Paper.

Kalish, S. / Nelson, P. (1991): A Comparison of Ranking, Rating and Reservation Price Measurement in Conjoint Analysis, in: *Marketing Letters*, 2:4, S. 327-335.

Köcher, W. (1997): Die MaiK-Conjoint-Analyse, in: *Marketing Zeitschrift für Forschung und Praxis*, 3, S. 141-152.

Krieger, A.M. / Green P.E. (1991): Designing Pareto Optimal Stimuli for Multiattribute Choice Experiments, in: *Marketing Letters*, 2:4, S. 337-348.

Kroeber-Riel, W. (1992): *Konsumentenverhalten*, 5. Aufl., München.

Kuhfeld, W.F. / Tobias, R.D. / Garratt, M. (1994): Efficient Experimental Design with Marketing Research Applications, in: *Journal of Marketing Research*, 31, S. 545-557.

Leigh, T.W. / MacKay, D.B. / Summers, J.O. (1984): Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison, in: *Journal of Marketing Research*, 21 (November), S. 456-462.

Levin, I.P. / Louviere, J.J. / Schepanski, A.A. / Norman, K.L. (1983): External Validity Tests of Laboratory Studies of Information Integration, in: *Organizational Behavior and Human Performance*, 31, S. 173-193.

Loosschilder, G.H. / Rosbergen, E. / Vriens, M. / Wittink, D.R. (1995): Pictorial Stimuli in Conjoint Analysis - to Support Product Styling Decisions, in: *Journal of Market Research Society*, 37 (1), S. 17-34.

Louviere, J.J. (1988): Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity, in: *Journal of Transport economics and Policy*, (January), S. 93-117.

Louviere, J.J. / Fox, M.F. / Moore, W.L. (1993): Cross-Task Validity Comparison of Stated Preference Choice Models, in: *Marketing Letters*, 4:3, S. 205-213.

Louviere, J.J. / Gaeth, G.J. (1988): A Comparison of Rating and Choice Responses in Conjoint Tasks, in: *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, ed. Ketchum, ID: Sawtooth Software, Inc., S. 59-73.

Louviere, J.J. / Woodworth, G. (1983): Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data, in: *Journal of Marketing Research*, 20 (November), S. 350-367.

Mahajan, V. / Green P.E. / Goldberg, S.M. (1982): A Conjoint Model for Measuring Self- and Cross-Price/Demand Relationships, in: *Journal of Marketing Research*, 19, S. 334-342.

Malhotra, N.K. (1986): An Approach to the Measurement of Consumer Preferences Using Limited Information, in: *Journal of Marketing Research*, 23 (February), S. 33-40.

Mehta, R. / Moore ,W. / Pavia, T.M. (1992): An Examination of the Use of Unacceptable Levels in Conjoint Analysis, in: *Journal of Consumer Research*, 19, S. 470-476.

Mengen, A. / Simon, H. (1996): Produkt- und Preisgestaltung mit Conjoint Measurement, in: WISU(3), S. 229-236.

Metegrano, M. (1994): *ACA System Version 4.0*, Evanston IL: Sawtooth Software, Inc.

Moore, R. (1992), "Doing Conjoint Analysis on the Telephone.", in: *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, ed. Ketchum, ID: Sawtooth Software, Inc., S. 245-255.

Moore, W.L. / Gray-Lee, J. / Louviere, J.J. (1998): A Cross-Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation, in: *Marketing Letters*, 9:2, S. 195-207.

Müller, S. / Kesselmann P. (1996), Akzeptanz von computergestützten Erhebungsverfahren: Ein empirischer Vergleich mit der traditionellen Fragebogentechnik, in: *Marketing Zeitschrift für Forschung und Praxis*, 3, S. 191-202.

Müller-Hagedorn, L. / Sewing, E. / Toporowski W. (1993): Zur Validität von Conjoint-Analysen, in: *Zeitschrift für betriebswirtschaftliche Forschung*, 45 (2), S. 123-148.

Neslin, S.A. (1981): Linking Product Features to Perceptions: Self-Stated Versus Statistically Revealed Importance Weights, in: *Journal of Marketing Research*, 18 (February), S. 80-86.

Oliphant, K. / Eagle, T.C. / Louviere, J.J. / Anderson, D. (1992): Cross-Task Comparison of Ratings-Based Choice-Based Conjoint, in: *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, ed. Ketchum, ID: Sawtooth Software, Inc., S. 383-404.

Oppedijk van Veen, W. M. / Beazley, D. (1977): An Investigation of Alternative Methods of Applying the Trade-off Model, in: *Journal of the Market Research Society*, 19 (1), S. 2-9.

Orme, B.K. / Alpert, M.I. / Christensen, E. (1997): Assessing the Validity of Conjoint Analysis - Continued, in: *Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, ed. Ketchum, ID: Sawtooth Software, Inc., S. 209-225.

Perry, J. (1996): Erhebungsdesign-Effekte bei der Conjoint-Analyse, in: *Marketing Zeitschrift für Forschung & Praxis*, 2 (2), S. 105-116.

Pullmann, M.E. / Dodson, K.J. / Moore, W.L. (1999): A Comparison of Conjoint Methods When There Are Many Attributes, in: *Marketing Letters*, 10:2, S. 125-138.

Rao, V.R. / Sattler, H. (1998): Measurement of Informational and Allocative Effects of Price, Working Paper, Johnson Graduate School of Management, Cornell University.

Rao, V. R. / Solgaard, H.S. (1978): An Empirical Evaluation of Alternative Multiattribute Utility Models, in: *Research Frontiers in Marketing: Dialoges and Directions*, 43, S. 27-31.

Reibstein, D. / Bateson, J.E.G. / Boulding, W. (1988): Conjoint Analysis Reliability: Empirical Findings, in: *Marketing Science*, 7 (3), S. 271-286.

Reiners, W. (1996): *Multiattributive Präferenzstrukturmodellierung durch die Conjoint Analyse: Diskussion der Verfahrensmöglichkeiten und Optimierung von Paarvergleichsaufgaben bei der adaptiven Conjoint Analyse*, Münster.

Reynolds, T.J. / Jolly, J.P. (1980): Measuring Personal Values: An Evaluation of Alternative Methods, in: *Journal of Marketing Research*, 17 (November), S. 531-536.

Rosko, M.D. / DeVita, M. / McKenna, W.F. / Walker, L.R. (1985): Strategic Marketing Applications of Conjoint Analysis: An HMO Perspective, in: *Journal of Health Care Marketing,* 5 (4), S. 27-38.

Rosko, M.D. / McKenna, W. (1983): Modelling Consumer Choices of Health Plans: a Comparison of Two Techniques, in: *Social Science Medicine*, 17 (7), S. 421-429.

Safizadeh, M. H. (1989): The Internal Validity of the Trade-off Method of Conjoint Analysis, in: *Decision Science*, 20, S. 451-461.

Sattler, H. (1991): *Herkunfts- und Gütezeichen im Kaufentscheidungsprozeß. Die Conjoint-Analyse als Instrument der Bedeutungsmessung*, Stuttgart.

Sattler, H. (1994): Die Validität von Produkttests. Ein empirischer Vergleich zwischen hypothetischer und realer Produktpräsentation, in: *Marketing Zeitschrift für Forschung und Praxis*, 16, S. 31-41.

Sattler, H. / Hensel-Börner, S. (2000): A Comparison of Conjoint Measurement with Self-Explicated Approaches, in: A. Gustafsson, A. Herrmann und F. Huber (Hrsg.): *Conjoint Measurement: Methods and Applications*, Berlin et al., S. 121-133.

Sattler, H. / Gedenk, K. / Hensel-Börner, S. (2000): Bandbreiten-Effekte bei multiattributiven Entscheidungen: Ein empirischer Vergleich von alternativen Verfahren zur Bestimmung von Eigenschaftsgewichten, Diskussionspapier der Wirtschaftswissenschaftlichen Fakultät der Friedrich-Schiller-Universität Jena.

Sattler, H. / Hensel-Börner, S. / Krüger, B. (2000): Die Abhängigkeit der Validität von Conjoint-Studien von demografischen Probanden-Charakteristika: Neue empirische Befunde, erscheint in: *Zeitschrift für Betriebswirtschaft*, 70. Jg.

Sattler, H. / Rao, V. (1997): Die Validität eines Ansatzes zur Separierung der Allokations- und Informationsfunktion des Preises, in: *Zeitschrift für Betriebswirtschaft*, 67. Jg. (12), S. 1285-1307.

Sawtooth Software (1995), The CBC System for Choice-Based Conjoint Analysis, URL: *http://www.skim.nl/Ssd_Tech.html#cbctech:* Sawtooth Software, Inc.

Sawtooth Software (1996), ACA System Adaptive Conjoint Analysis Version 4.0, URL: *http://www.sawtoothsoftware.com/TechPap.htm:* Sawtooth Software, Inc.

Schubert, B. (1995): Conjoint-Analyse, in: *Handwörterbuch für Marketing, Tietz B., R. Köhler und J. Zentes (Hrsg.)*, Stuttgart, S. 376-389.

Schweikl, H. (1985): *Computergestützte Präferenzanlayse mit individuell wichtigsten Produktmerkmalen*, Berlin.

Scott, J.E. / Keiser, S.K. (1984): Forecasting Acceptance of New Industrial Products with Judgement Modeling, in: *Journal of Marketing*, 84 (Spring), S. 54-67.

Segal, M.N. (1982): Reliability of Conjoint Analysis: Contrasting Data Collection Procedures, in: *Journal of Marketing Research*, 19 (February), S. 139-143.

Silk, A.J. / Urban G.L. (1978): Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology, in: *Journal of Marketing Research*, 15, S. 171-191.

Srinivasan, V. (1988): A Conjunctive-compensatory Approach to the Self-Explication of Multiattributed Preferences, in: *Decision Science*, 19 (Spring), S. 295-305.

Srinivasan, V. / deMaCarty, P. (1998): An Alternative Approach to the Predictive Validation of Conjoint Models, Research Paper No. 1483 (March), Graduate School of Business, Standford University.

Srinivasan, V. / Jain, A.K. / Malhotra, N.K. (1983): Improving Predictive Power of Conjoint Analysis by Constrained Parameter Estimation, in: *Journal of Marketing Research*, 10 (November), S. 433-438.

Srinivasan, V. / Park, C.S. (1997): Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement, in: *Journal of Marketing Research*, 34, S. 286-291.

Srinivasan, V. / Shocker, A.D. (1973): Linear Programming Techniques for Multidimensional Analysis of Preferences, in: *Psychometrika*, 38 (3), S. 337-369.

Srinivasan, V. / Wyner, G.A. (1989): CASEMAP: Computer-Assisted Self-Explication of Multiattribute Preferences, in: *New Product Development and Testing, W. Henry, M. Menasco und H. Takada (Hrsg.),* Lexington, Mass., S. 91-111.

Stallmeier, C. (1993): *Die Bedeutung der Datenerhebungsmethode und des Untersuchungsdesigns für die Ergebnisstabilität der Conjoint-Analyse*, Regensburg.

Steenkamp, J.B. / Wittink, D.R. (1994): The Metric Quality of Full-profile Judgments and the Number-of-Attribute-Levels Effect in Conjoint Analysis, in: *International Journal of Research in Marketing*, 11, S. 275-286.

Teas, K. / Dellva, W.L. (1985): Conjoint Measurement of Consumers' Preferences for Multiattribute Financial Services, in: *Journal of Bank Research* (Summer), S. 99-112.

Thomas, L. (1979): Conjoint Measurement als Instrument der Absatzforschung, in: *Marketing Zeitschrift für Forschung und Praxis*, 3, S. 199-211.

van der Lans, I.A. / Heiser, W.J. (1992): Constrained Part-Worth Estimation in Conjoint Analysis Using Self-explicated Utility Model, in: *International Journal of Research in Marketing*, 9, S. 325-344.

Voeth, M. (1999): 25 Jahre conjointanalytische Forschung in Deutschland, in: *ZfB-Ergänzungsheft*, Heft 2, S. 153-176.

Voeth, M. / Hahn, C. (1998): Limit Conjoint-Analyse, in: *Marketing Zeitschrift für Forschung und Praxis*, 2, S. 119-132.

von Nitzsch, R. / Weber, M. (1991): Bandbreiten-Effekte bei der Bestimmung von Zielgewichten, in: *Zeitschrift für betriebswirtschaftliche Forschung*, 43, S. 971-985.

Vriens, M. (1995): *Conjoint Analysis in Marketing: Developments in Stimulus Representation and Segmentation Methods*, Capelle a/d Ijssel.

Weiber, R. / Rosendahl, T. (1997): Anwendungsprobleme der Conjoint-Analyse. Die Eignung conjointanalytischer Untersuchungsansätze zur Abbildung realer Entscheidungsprozesse, in: *Marketing Zeitschrift für Forschung und Praxis*, (2), S. 107-118.

Wilkie, W. / Pessemier, E. (1973): Issues in Marketing's Use of Multi-Attribute Attitude Models, in: *Journal of Marketing Research*, 10 (November), S. 428-441.

Wittink, D.R. / Bergestuen, T. (1999): Forecasting With Conjoint Analysis, erscheint in: J. Scott Armstrong (ed.): *Principles of Forecasting: A Handbook for Researchers and Practioniers,* Norwell, MA: Kluwer Academic Publisher, S. 1-19.

Wittink, D.R. / Cattin, P. (1981): Alternative Estimation Methods for Conjoint Analysis: a Monte Carlo Study, in: *Journal of Marketing Research*, 23 (February), S. 101-106.

Wittink, D.R. / Cattin, P. (1989): Commercial Use of Conjoint Analysis: an Update.", in: *Journal of Marketing*, 53 (July), S. 91-96.

Wittink, D R. / Krishnamurthi, L. / Nutter, J.B. (1982): Comparing Derived Importance Weights Across Attributes, in: *Journal of Consumer Research*, 8, S. 471-474.

Wittink, D.R. / Krishnamurthi, L. / Reibstein, D.J. (1989): The Effect of Differences in the Number of Attribute Levels on Conjoint Results, in: *Marketing Letters*, 1:2, S. 113-123.

Wittink, D.R. / Vriens, M. / Burhenne, W. (1994): Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections, in: *International Journal of Research in Marketing*, 11 (1), S. 41-42.

Wright, P. / Kriewall, M.A. (1980): State-of-Mind Effects on the Accuracy with Which Utility Functions Predict Marketplace Choice, in: *Journal of Marketing Research*, 17 (August), S. 277-293.

Zwerina, K. / Huber, J. (1996): Deriving Individual Preference Structures from Practical Choice-Experiments, Working Paper, Duke University, August.